ANALYSIS

Edited by
CHRISTOPHER KIRWAN

(7)

CONTENTS

105

NAMES, INDICES AND INDIVIDUALS　　An 13

By William Godfrey-Smith

IN this paper I will extend Geach's account of an 'act of naming',[1] which I will suggest can help to elucidate Kripke's account of names as *rigid designators*.[2] The importance of acts of naming for our understanding of names is that it is through such acts that names acquire the elusive indexical character which Russell, for example, was able to locate only in the case of demonstratives; and which led him to an acquaintance theory of reference.[3] Russell seems ultimately to have been driven to the conclusion that the only items to which one can directly refer are objects of one's immediate experience—sense data. This conclusion, I suggest, is untenable; and Kripke's account of names as genuine indices —which I will argue for by comparing it with the account of names proposed by Peirce[4]—provides an effective way of avoiding Russell's conclusion. It is through the use of names in acts of naming that these expressions come to designate rigidly. I shall go on to suggest that the notion of a rigid designator is helpful in turn for coming to grips with the basic notion of an individual.

I

Frege held a description theory of names, which he seems to have thought was needed to explain the contribution which names make to the truth of propositions in which they occur. An important feature of names, however, noted by Geach, is that they occur in contexts which do not involve the use of assertoric propositions, and in such cases questions of truth do not arise. Geach calls such uses 'acts of naming' (op. cit., p. 26), and they are important for our understanding of names. The first example of an act of naming which Geach provides is what I shall call an 'act of greeting', in which a name is used to acknowledge the presence of an individual. We may say, for example, 'Hullo, Jemima!' to acknowledge the presence of an individual cat. Acknowledgments do not serve to *describe* an individual in any way; though we could not of course identify an individual if we did not know what sort of individual it is. It is important to note that it cannot be solely on the basis

---

[1] P. T. Geach, *Reference and Generality* (Ithaca: Cornell University Press, emended ed. 1968), §§ 20, 32.

[2] S. Kripke, 'Naming and Necessity', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language* (Dordrecht: Reidel, 1972), pp. 253–355.

[3] B. A. W. Russell, 'Knowledge by Acquaintance and Knowledge by Description', *Proceedings of the Aristotelian Society* XI (1910–11), pp. 108–128, reprinted in *Mysticism and Logic* (London: Longmans Green & Co, 1918), pp. 209–232.

[4] Cf. C. S. Peirce, *Collected Papers*, C. Hartshorne and P. Weiss (eds.), 6 Vols. (Cambridge, Mass.: Harvard University Press, 1931–35), 2.287, 2.329, 3.419.

of any characterization of an individual which we greet that the name
secures its reference. For no amount of description could serve to
distinguish *that* individual from other (actual or possible) individuals of
that sort. What needs to be explained, then, is how we refer to a parti-
cular individual, *this* cat, say, rather than to *an* individual of a particular
sort.

Acts of greeting provide a useful criterion for namehood. Geach
also employs the 'act of naming' in his account of demonstrative refer-
ence, though this case differs from the act of greeting in the respect that
it need not employ a proper name at all. For example, 'That man nearly
got sent to prison', according to Geach, incorporates an act of naming
and a predication (ibid., pp. 40, f.). In contrast to acts of greeting, how-
ever, there is no need in this sort of case to employ proper names at all;
and this of course is one of the great utilities of demonstrative reference.
It enables us to pick out and talk about an individual whose name we may
not happen to know, or in whom our interest is not sufficiently enduring
to justify finding its name, or even bestowing a name on it at all.
Zoological and botanical specimens are frequently treated in this way.
The act of greeting use also shows that demonstratives are not names;
it seems ludicrous to suggest that we could acknowledge the presence
of an individual by saying 'Hullo, this!' or 'Hullo, this cat!'

There are two further cases which also seem to be naturally charac-
terized as acts of naming of a sort. One is the act of *christening* in which
a name is actually bestowed on an individual. We might employ

   (1) This cat is Jemima

to confer a name on an individual cat; though it might be felt that a
more ceremonious performative formula would be more appropriate.
The ceremonial aspects however are clearly of no *logical* interest. More
commonly (1) would be employed in acts of *introduction,* which do not
serve to confer a name, but to inform someone of the name of an in-
dividual with which he or she was not previously acquainted. In an act
of introduction the force of (1) is: 'This cat is *called* Jemima'. In this case
the name is treated as a property of the individual and it does not involve
the *use* of the name; though introductions are commonly preliminaries
to the use of a name, just as christenings are commonly preliminaries to
introductions.

In the case of an introduction there is always the possibility that the
individual may have been misidentified by the introducer, so it may be
that (1) is false. This shows that (1) differs in such uses from acts of
greeting since it is clearly propositional. Acts of christening, in contrast,
fix the use of a name in a way which precludes this sort of mistake;
though there are other ways in which the individual thus named may
be misidentified. It might be believed that this individual is the φer,

and that may not be so. Acts of christening are not propositional at all; they are of course one of the principal instances of nonassertoric uses of a sentence identified by Austin and characterized by him as *performatives*.[1]

Names are usually introduced through the use of sentences which employ demonstratives. The way in which we establish or fix the use of a name standardly employs demonstrative reference to an individual, as in an act of christening. The demonstrative introduction of a name can be remote from our contemporary use of it, but if the expression is a genuine name there is a chain of reference reaching back to such a use. This is a basic claim of Kripke's account of names which I will consider in part II below.

Geach treats the demonstrative 'this' as a particularizing prefix which transforms a substantival general term into a name. Thus 'this cat' can be used to pick out an individual which is named by the common noun 'cat'. This approach correctly avoids treating 'this' as a name. Such phrases as 'this cat'—which I shall call a 'demonstrative referring complex'—involve our immediate attention to an individual in a sensory context in which the individual is directly *indicated*; and this distinguishes demonstrative referring complexes from proper names. However a demonstrative referring complex can be attached to a predicable in exactly the same way that a name can. It seems that

(2) Jemima is waiting for a mouse,

in appropriate circumstances, is equivalent to

(3) This cat is waiting for a mouse.

In fact (3) is apparently less subject to error than (2), for it may be that the speaker has misidentified the cat in question. However it would still be the case that (3) is true so long as some cat in the appropriate sensory context is waiting for a mouse. Demonstrative referring complexes in this way can override the use of a name. This overriding use also distinguishes demonstrative indication from naming.

It seems clear that, if (3) is true, then, so long as (1) is correct, (2) will follow. It might be thought because of this that (1) in some cases is a statement of identity; but I think this would be a mistake. Acts of introduction involve the identification of an individual, and because the demonstrative referring complex overrides the name 'Jemima' this latter expression is not being used referentially. Clearly (1) is different in force from

(4) Jemima is Jemima;

and also different from

(5) This cat is this cat.

---

[1] J. L. Austin, *How to do Things with Words* (Oxford: Oxford University Press, 1962).

Certainly (4), and—if it says anything—(5), are statements of identity. Propositions like (1) are statements of *identification* when they serve to introduce names, and it is the possibility of being mistaken which suggests that in such uses (1) really amounts to 'This cat is *called* Jemima'. Propositions such as (3) are better characterized as ones which are used in an act of indication rather than an act of naming; so as to distinguish them from acts of christening, greeting and introduction; all of which must employ a proper name.

It is through the use of sentences such as (1) employed in acts of christening that names are attached to their bearers; and their use in acts of introduction is a common way in which we learn which individual a name denotes. Once attached in this way names continue to designate *the same* individual. Neither naming an individual nor indicating it are ways of describing it: naming is a preparation for describing. Though *having* a particular name is a property of an individual, it is a mistake to regard a name as something *true of* its bearer. There is a tendency to think that because a name is contingently attached to its bearer, and might have been used to designate another individual, names therefore pick out their bearers accidentally. We must, I think, attend carefully to the distinction between *being* N and *being called* N if we are to avoid falling into serious error. This is a distinction which Russell was quite clear about in the following passage:

> A proposition containing a description is not identical with what that proposition becomes when a name is substituted, even if the name names the same object as the description describes. 'Scott is the author of *Waverley*' is obviously a different proposition from 'Scott is Scott': the first is a fact of literary history, the second a trivial truism. . . . But, it may be said, our proposition is essentially of the same form as (say) 'Scott is Sir Walter', in which case two names are said to apply to the same person. The reply is that, if 'Scott is Sir Walter' really means 'the person named "Scott" is the person named "Sir Walter"', then the names are being used as descriptions: *i.e.* the individual, instead of being named, is being described as the person having that name. This is a way in which names are frequently used in practice, and there will, as a rule, be nothing in the phraseology to show whether they are being used in this way or *as* names.[1]

Russell goes on to point out that names used *as* names are logically structureless: 'When a name is used directly, merely to indicate what we are speaking about, it is no part of the *fact* asserted' (ibid.). Russell here actually appears to be uncharacteristically siding with Mill rather than Frege about names; but perhaps if pressed he would have said that the expression 'this' is the only candidate which really qualifies as a logically proper name.[2]

[1] Russell, *Introduction to Mathematical Philosophy* (London: Allen & Unwin, 1919), pp. 174f.
[2] Cf. Russell, 'Knowledge by Acquaintance and Knowledge by Description', p. 224.

Indicating may be a preparation for describing, or it may be an act performed simply to draw our attention to the presence of an individual. It is not a way of *naming* an individual, since for a name to be a name of an individual we must presuppose some (actual or possible) act of indicating that individual through the use of a demonstrative referring complex. Indicating is an activity distinct from naming (and describing) and indicating—or the possibility of indicating—is essential for fixing the reference of a name. For an expression to be a name it must answer back to an act of christening, or at least the possibility of such an act; and it should have been possible at some time to have employed the name in an act of greeting to acknowledge the presence of the individual designated by the name.

II

The central claim of the account of names proposed by Kripke is that they are *rigid designators,* and it is worth examining the way in which the notion of a rigid designator is introduced. Kripke begins by discussing *Nixon,* designated as 'the man who won the election in 1968' (op. cit., p. 265). 'When you ask whether it is necessary or contingent that *Nixon* won the election, you are asking the intuitive question whether in some counterfactual situation *this man* would in fact have lost the election' (ibid.). And 'the term "Nixon" is just a *name* of *this man*' (ibid.). What is it that makes Nixon *this man*? Kripke claims that we can consider counterfactual situations in which we change various properties of Nixon, but these possible worlds are ones which contain *this* man. This is a situation which we do not need to describe, and indeed ought not to describe, as one which contains an individual who qualitatively *resembles* Nixon; the situation

> need not be identified with the possibility of a man looking like such and and such, or holding such and such political views, or otherwise qualitatively described . . . We can point to the *man,* and ask what might have happened to *him,* had events been different. (Ibid., p. 268.)

Kripke then introduces a more immediate example—a table which he directly perceives—and asks whether *it* might have been different. It is clear here that Kripke is relying on demonstratives to try and bring out a point about individuals to which he refers. But the only way in which he seems to be able to make the point is by emphasizing pronouns, demonstratives and phrases like *'this very thing'.* One might be disposed here to ask: *Which* very thing? And the only means by which we seem to be able to say *what* Kripke is talking about is by providing some description of the person or object to which he is referring. Kripke's contention however is that no specification of properties will serve to explain what makes this object *this very thing.*

The table is a particular. Can we provide a criterion of identity for it somehow? Kripke allows that we can: we can provide a criterion in terms of other particulars, namely its constituent parts. But if

> it is demanded that I describe each counterfactual situation purely qualitatively, then I can only ask whether *a table*, of such and such a color, and so on, would have certain properties; whether the table in question would be *this table*, table *T*, is indeed moot, since all reference to objects, as opposed to qualities, has disappeared. (Ibid., p. 272.)

Kripke clearly believes that the Leibnizian account of objects as collections of properties is mistaken. But it seems that the alternative to this is to posit some mysterious bare particular or Lockean propertyless substratum as that which underlies the qualities. However according to Kripke this is a false dilemma:

> Philosophers have asked, are these objects *behind* the bundle of qualities, or is the object *nothing but* the bundle? Neither is the case; this table is wooden, brown, in the room, etc. It has all these properties and is not a thing without properties, behind them; but it should not be identified with the set or 'bundle' of its properties, nor with the subset of its essential properties. Don't ask: how can I identify this table in another possible world, except by its properties? (Ibid.)

It is all very well to be told *not* to ask, but I think one may be forgiven for *wanting* an answer to just this question. How indeed do I identify the table in *this* world except by its properties? Kripke continues: 'I have the table in my hands, I can point to it, and when I ask whether *it* might have been in another room, I am talking, by definition, about *it*' (ibid., pp. 272f). But this does not seem to remove the puzzlement: *what*, it may be wondered, *is* Kripke talking about?

It seems to me nevertheless that Kripke *has* succeeded in latching on to the table, and in this passage he is trying to bring out a point of some importance. Kripke says he has the table in his hands. He has quite unquestionably and literally got hold of it. But no amount of specification of properties can tell us what he has done. Kripke continues:

> properties are not used to identify the object in another possible world for such identification is not needed; nor need the essential properties of an object be the properties used to identify it in the actual world, if indeed it is identified in the actual world by means of properties. (Ibid., p. 273.)

Kripke says he wants to leave this question open. However the general drift of his remarks suggests that he does not think that the object can be identified solely by means of some specification of its properties; but the matter is not pursued. I think we can obtain further clarification of the problematic notion of an individual from an account put forward by Vendler. Vendler, I think, develops further the point hinted at here by

Kripke, which is that our grasp of particulars is something which in the end has to be taken quite literally. The account has other affinities with Kripke's views as well. Vendler writes:

> Descriptions, I believe, achieve reference to spatiotemporal individuals only insofar as they form a chain, actually given or presupposed in the discourse, leading to some 'basic' individuals which have to be named or have to be pointed at in an ostensive situation. It seems to me, moreover, that even the use of names is effective only because a similar chain connects them with things encountered in ostensive situations.[1]

The problem, as Vendler puts it, is 'what in thought corresponds to ostension and naming?' (ibid.). That is, how is it that a concrete particular can be grasped as an object of thought? The answer which Vendler gives is that it cannot:

> Descriptions, insofar as they do not contain names and indexicals, can be understood; names and indexicals, however, cannot. As the mediaevals said, there is no understanding of the individual; no concept can capture its *haecceitas*. We understand the 'such', not the 'this', and no amount of 'such' congeals into a 'this' . . . What it is for something to be a book, to be old, to be bound in leather, to be lying on the table, and so on, are perfectly understandable ideas. What is not a matter for the understanding, however, is that these attributes are *de facto* realized, and realized in this one subject . . . That they *de facto* are realized at all, and realized in this subject, is a matter for the senses and not for the understanding.
> (Ibid., pp. 74f.)

It should be noted here that Vendler assimilates names to the category of indexical expressions rather than to that of descriptive expressions. This I think is a central point which emerges from Kripke's account of names. It also shows that Locke's point about the unintelligibility of substance has a sound basis.[2] Where the Lockean account goes wrong is in supposing that the individual is somehow "behind" the properties; clearly suggesting that *it* is something which is itself without properties. Kripke's rejection of the dilemma is sound, and Vendler helps to bring out where the mistake lies. When I point to an individual I am not pointing to a collection of properties, but to a thing with properties which can, at least in paradigm instances, be quite literally grasped. However the particularity of the thing is not a matter for the intellect. Thought is essentially general and cannot, of itself, provide the wherewithal to grasp particulars.[3] 'The real world', Peirce remarked, 'cannot be distinguished from a fictitious world by any description' (op. cit., 2. 337).

The claim that individuals are unintelligible; that is, that concrete particulars cannot be grasped through the understanding by any

[1] Z. Vendler, *Res Cogitans* (Ithaca: Cornell University Press, 1972), p. 73.
[2] Locke, *An Essay Concerning Human Understanding* (London, 1690), Bk. II, Ch. 23.
[3] Cf. Geach, *Mental Acts* (London: Routledge & Kegan Paul, 1957), p. 65.

amount of description, comes out very clearly in the writings of Peirce. Peirce contrasted the 'diagrammatic'—roughly, descriptive—with the indexical, which involves a direct causal effect on the senses:

> Diagrams and diagrammatoidal figures are intended to be applied to the better understanding of states of things . . . Such a figure cannot, however, show what it is to which it is intended to be applied; nor can any other diagram avail for that purpose. The where and the when of the particular experience . . . to which the diagram is to be applied, are things not capable of being diagrammatically exhibited. Describe and describe and describe, and you can never describe a date, a position, or any homaloidal quantity. You may object that a map is a diagram showing localities; undoubtedly, but not until the law of the projection is understood, and not even then unless at least two points on the map are somehow previously identified with points in nature. Now, how is any diagram ever to perform that identification? If a diagram cannot do it, algebra cannot: for algebra is but a sort of diagram; and if algebra cannot do it, language cannot: for language is but a kind of algebra. It would, certainly, in one sense be extravagant to say that we can never tell what we are talking about; yet, in another sense, it is quite true. (Ibid., 3. 419.)

What we grasp indexically, then, is not something for the understanding. It is in this sense that we can never tell—that is, grasp intellectually— what the "it" is that we are talking about. The function of indexicals is to surmount this limitation of purely descriptive language by placing us, as it were, in direct contact with reality. Peirce continues:

> Words like *this, that, lo, hallo, hi there,* have a direct, foreful action upon the nervous system, and compel the hearer to look about him; and so they, more than ordinary words, contribute towards indicating what the speech is about. But this is a point that grammar and the grammarians . . . are so far from seeing as to call demonstratives, such as *that* and *this,* pronouns—a literally preposterous designation, for nouns may be more truly called pro-demonstratives. (Ibid.)

Peirce's suggestion here seems to come close to the Russellian view that demonstratives are the only genuine indexical expressions, and this, I think, is a mistake. Names are *introduced by* demonstratives, but it does not follow that names stand *in place of* demonstratives. This should be obvious from the fact that demonstratives can be used to pick out individuals only in an appropriate sensory context, whereas in the case of names there is no such requirement; and it is important for their linguistic function that there should not be. Peirce's suggestion that names *do* stand in place of demonstratives seems to be further reinforced in the following passage:

> There is no reason for saying that *I, thou, that, this,* stand in place of nouns; they indicate things in the directest possible way. It is impossible to express what an assertion refers to except by means of an index. A pronoun is an index. A noun, on the other hand, does not *indicate* the object it denotes; and when a noun is used to show what one is talking about, the experience of the hearer is relied upon to make up for the incapacity

of the noun for doing what the pronoun does at once. Thus a noun is an imperfect substitute for a pronoun ... Allen and Greenough say, 'pronouns indicate some person or thing without either naming or describing' [*New Latin Grammar* (1884), p. 128]. This is correct—refreshingly correct; only it seems better to say what they *do,* and not merely what they don't. (Ibid., 2. 287n.)

Peirce's claim that demonstratives cannot be assimilated to the semantic category of names or to that of descriptive expressions is, I think, correct. What demonstratives *do* is to *indicate,* and indicating is neither naming nor describing. But his suggestion that names are 'imperfect substitutes' for pronouns is, on the face of it, mistaken. Nevertheless the dependency here does run in the right direction: indicating is not to be assimilated to naming, for naming is an indirect form of indicating. The indirect indication, Peirce said, has to be mediated by the experience of the hearer. This I think need not be interpreted as the Fregean claim that the reference of a name is determined by its sense. It is not part of Peirce's claim that names have senses: as he put it, 'there is no verb wrapped up in a proper name' (ibid., 2. 328). This seems to embody the view of Mill, defended by Kripke against Russell and Frege, that proper names denote without connoting. Names on this view are pure 'indices'; but Peirce goes on to say that this is true of a proper name only 'when one meets it for the first time', when it 'is existentially connected with some percept or other equivalent individual knowledge of the individual it names' (ibid., 2. 329). Subsequently the name is merely 'an Icon of that Index' (ibid.). It seems to me that here Peirce goes wrong by adopting the mistaken Russellian view that demonstrative reference does, after all, provide us with the only genuine indices. And if this is granted then it is natural to suppose that in all other cases names are really disguised descriptions.

Kripke's suggestion that causally-based reference-preserving links can sustain names as genuine indices is important if we are to avoid the path taken by Russell. I think that in the final analysis it was dubious Cartesian epistemology which lay behind Russell's mistaken view about names. In particular I think it must be rejected because it places unacceptable and counterintuitive restrictions on the individuals which we can pick out and talk about. In order to be able to use a proper name correctly we do need some background knowledge about the individual named—indeed it would be hard to see what the *point* of naming might be if such information was lacking—but this may be regarded as part of the context in which the name occurs. However these contextual features are irrelevant to the *logical* role played by names in designating their bearers: they do not constitute a logical structure *of* the name. There is an epistemological ingredient which must inevitably be considered in connection with the use of a name, but this has no bearing on the logical

role played by names in picking out their bearers.[1] We should also note, as Kripke has pointed out, that the background knowledge needed for our use of names need not be uniquely individuating.

Australian National University   © WILLIAM GODFREY-SMITH 1976

    [1] Cf. A. W. Müller, 'Reply to A. N. Prior's "I" ', in B. Y. Khanbhai *et al.* (eds.), *Jowett Papers*, 1968–69 (Oxford: Blackwell, 1970), p. 14.

THE ALLEGED PARADOX OF DEMOCRACY

By VINIT HAKSAR

B RIAN BARRY says 'If you adhere to any ideal or principle which does not include in it a reference to the opinion of others then it is logically possible that you might be the only person holding it. There could be a situation, therefore, in which you say "So and so should be done" and everyone except you says "So and so should not be done". You are in effect setting yourself up as a dictator. Of course, you probably won't have the power to get what you want done against everyone's opposition, nevertheless you are saying that if you had the power you would [do it]. (You cannot sincerely say "So and so should be done" and then not do it if you had the power.)'[1]

Richard Wollheim[2] describes a similar situation which gives rise to what he calls the paradox of democracy. Suppose you say 'A ought to be done' and suppose you are a democrat and the democratic (i.e. majority) decision is that B ought to be done; and suppose that A and B are incompatible: then you are committed both to the view that A ought to be done and to the view that A ought not to be done.

It seems to me that such paradoxes disappear once we appreciate the significance of certain distinctions. We need to distinguish the following: (1) Ought A to be done? and (2) If so, which means of achieving A are legitimate? Who, if anyone, has the relevant authority to do A? I might believe that wealth ought to be redistributed from X to Y. According to Barry (in the passage I have quoted from him) this commits me to the view that if I had the power, I should carry out the redistribution. But this is not so. There is at least one missing premiss, viz. I have the relevant authority. Suppose I have the power to steal money from A and give it to B and get away with the theft: it does not

    [1] Brian Barry, *Political Argument*, Routledge and Kegan Paul, 1965, p. 59.
    [2] R. Wollheim, 'A Paradox in the Theory of Democracy', *Philosophy, Politics and Society*, Second Series, eds. P. Laslett and W. G. Runciman, Blackwell, 1962.

follow that I should carry out the theft. If I have no authority to re-
distribute the money, then it does not follow that I should take the
money from A and give it to B, even if it is the case that money ought
to be transferred from A to B. If I think an end ought to be achieved,
this surely does not commit me to the view that any means, however
illegitimate, towards that end would be justified. The most that follows
is that (morally) legitimate means towards that end should be used.

Now to come back to democracy. If I believe that there ought to be
(compulsory) redistribution of wealth from the rich to the poor, this
commits me to the view that legitimate means should be used to imple-
ment this goal: it does not commit me to the view that illegitimate
means should be used. Now suppose I am a member of a democratic
parliament and propose that parliament should pass a law authorizing
and instructing the State to redistribute wealth from the rich to the
poor; and suppose there is a vote and my proposal is defeated. I could
still maintain, with all my old fervour, that there ought to be compulsory
redistribution of wealth from the rich to the poor; but this is quite
consistent with my admission (as a good democrat) that no one as yet
has the authority to carry out this redistribution.

The above points that I have made can be reinforced by an appeal to
the defeasible nature of practical reasoning pointed out by Kenny and
Geach.[1] The following instance of practical reasoning is defeasible:

P_1: There ought to be compulsory redistribution of wealth.
P_2: X has the power to carry out the compulsory redistribution of
    wealth.
T: Therefore X ought to carry out the compulsory redistribution
    of wealth.

The above reasoning is defeasible, for it can be invalidated by bringing
in an additional premiss such as P_3: X has no authority to carry out the
compulsory redistribution of wealth.

Now a democrat could believe that majority voting should decide
whether or not any person has the authority to carry out the redistri-
bution of wealth. So a democrat could maintain that the majority voting
will determine whether or not P_3 is a satisfactory premiss: if the majority
decides not to authorize X, P_3 is satisfactory, if the majority decides to
authorize X, P_3 is not satisfactory. Now if P_3 is a satisfactory premiss, it
can be used as an additional premiss (i.e. in addition to the original
premisses P_1 and P_2), and so can serve to invalidate the practical in-
ference from the premisses to T. In this way the majority decision can
be relevant to whether or not the practical inference is invalidated. A

[1] See A. Kenny, *Will, Freedom and Power*, Blackwell, 1975, pp. 92–96, and P. T. Geach,
'Dr Kenny on Practical Reasoning', ANALYSIS, 26.3, 1966, pp. 76–79, reprinted in P. T. Geach,
Logic Matters, pp. 285–288

person could be a democrat if he allows the majority decision to be a determinant of whether or not the relevant practical inference (from 'Such and such ought to be done' to 'So and so ought to do it') is to be invalidated. He does not, in order to be a democrat, have to believe that the majority decision is a determinant of any of the original premisses $(P_1$ and $P_2)$ of the practical inference. The view that the majority decision is a determinant of whether a practical inference is valid does not commit one to the view that the democratic decision is a determinant of any of the original premisses of the practical inference. So I can be a democrat without getting involved in paradoxes such as being simultaneously committed to the view that P_1 is right, because I believe it is right, and to the view that P_1 is not right, because the majority thinks it is not right. But, though one can be a democrat without being involved in this paradox, can one be a genuine democrat without being involved in it? (A genuine democrat is one who does not value democracy merely as a means to other goals.) I shall now argue that we can.

I think Wollheim's alleged paradox looks powerful because of a false dichotomy that he presupposes (op. cit., pp. 64, sq.): either a democrat is genuine, in which case he must agree that the majority is right about what ought to be done; or he is not a genuine democrat but just a tactical or prudential democrat, who accepts democratic results for tactical or prudential reasons. But Wollheim's dichotomy is a false one. A person can be a genuine democrat, in the sense that he is morally committed to democracy, and yet not agree that the majority is right about what ought to be done. He can be a genuine democrat because he genuinely (and not just tactically or prudentially) accepts and believes that the majority has the (moral) right to authorize or refuse to authorize the use of certain means and certain compulsory measures. This does not commit him to believing that the majority exercises such rights wisely.

Keith Graham[1] considers some views (e.g., those of R. Pennock[2]) similar to mine but rejects them because he too presupposes Wollheim's false dichotomy. Graham goes even further than Wollheim and insists that the genuine democrat must not only regard democracy as morally desirable, but must also regard democratic practice (such as voting) as morally desirable in itself. However, it seems to me that my solution is compatible with the view that democratic practice is morally desirable in itself. Those who don't think the majority is necessarily right with regard to what ought to be done, but insist on the use of democratic means (as opposed, say, to totalitarian means) are not necessarily treating democracy merely as a means to some goal of theirs. There is nothing

---

[1] K. Graham, 'Democracy, Paradox and the Real World', *Proc. of the Arist. Soc.*, 1975–76, pp. 231–232.
[2] R. Pennock, 'Democracy is not Paradoxical', *Political Theory*, Vol. 2, 1974, pp. 88–93.

absurd in the view that the use of such democratic means is morally
desirable in itself, just as there is nothing absurd in the view that the use
of totalitarian means is morally undesirable in itself.

Let us call the person who believes that the majority is always right
about what ought to be done, a majoritarian.[1] Now it may be objected
that, although I have shown that Wollheim's paradox does not arise
for all democrats, not even for all genuine democrats, yet the paradox
can arise for a sub-class of genuine democrats, i.e. for those genuine
democrats who are majoritarians. However, it may be replied that the
majoritarian can get out of the paradox in the manner suggested by
Barry. Barry's suggestion is that the majoritarian corrects his primary
wish 'A ought to be done' by the majority wish 'B ought to be done',
and so his corrected wish becomes 'B ought to be done'.

Barry's solution of the paradox has been criticized. Ross Harrison[2]
says that Barry cannot explain the fact that democrats often vote for
policies which they know in advance will be rejected by the majority.
Now I think Barry's solution can be defended against such criticisms.
I think such criticisms look powerful only if one presupposes the false
dichotomy that I referred to earlier. Once we realize that not all genuine
democrats are majoritarians, we can grant that democrats, even genuine
democrats, often vote for policies which they know in advance will be
rejected by the majority, and yet can doubt whether it makes sense for
majoritarians to vote for policies which they know in advance will be
rejected by the majority. I believe that in practice all rational (and moral)
democrats who vote for a policy knowing that it will be rejected by the
majority are non-majoritarian democrats. But it may be asked what hap-
pens if we come across a majoritarian who votes for a policy knowing
that it will be rejected by the majority. I am inclined to say that such a
majoritarian is being irrational (or immoral), and we should not sym-
pathize with the predicament that a cantankerous majoritarian has need-
lessly got involved in.

Ross Harrison, too, fails to distinguish the sub-class of majoritarians
from the wider class of genuine democrats, and so he too is sympathetic
to the predicament of the majoritarian. He compares the majoritarian's
plight to that of the person torn between different moral principles, and
he thinks the solution to the alleged paradox is the same as the solution
to the wider problem of what we should do when two moral principles
conflict with each other: 'He can decide, that is, to let one of his prin-
ciples override the other or else, perhaps, be torn tragically between
them' (ibid., p. 516). It seems to me that the sensible thing is to qualify
the majoritarian principle so that it does not conflict with the agent's
moral views. I agree that often in moral life we are faced with a genuine

---

[1] Cf. Barry, op. cit., Chapter 4.
[2] R. Harrison, 'No paradox in Democracy', *Political Studies*, 1970, pp. 514–517.

conflict and we may have to let one of the principles be overridden. But in some cases of alleged conflict the sensible thing is suitably to qualify one of the principles, so that it does not conflict with the other.[1] It seems to me that the most one can claim for the majoritarian principle is that it should apply to areas where matters of individual integrity are not at stake, where the answer to the question what ought to be done is fairly indeterminate; in such areas it may be quite sensible to take the line that we should follow and be converted by the majority judgment on what is right. But in other more important areas majoritarianism is immoral and ought to be abandoned, so that the problem of over-riding it won't arise. This conclusion should become palatable once we realize that to give up being a majoritarian in these important areas is not necessarily to give up being a genuine democrat in these important areas. So in some areas majoritarianism is immoral and ought to be abandoned; whereas in other areas, where majoritarianism could reasonably apply, there is again no paradox, because the sensible thing for the majoritarian to do in such relatively unimportant areas is to correct his primary wishes by the majoritarian wishes in the manner suggested by Barry.

My criticism of Harrison is quite different from Graham's criticism of Harrison. Graham says 'It just so happens that we occasionally encounter drowning men whilst on our way to keep promised appointments, and we can always do our best to ensure that we avoid such situations or hope that the world will be kinder. But in a system where decisions are recurrently taken on the basis of counting votes, every time a democrat is on the losing side (and someone has to be) he is faced with the dilemma: the problem is endemic.' (Op. cit., p. 233.) However, it seems to me that if majoritarianism is suitably modified in the way that I suggested earlier there will be no dilemma for the majoritarian who is on the losing side; at any rate, the problem will not be endemic. So while Graham's complaint is that Harrison is not sufficiently impressed by the predicament facing the democrat, my complaint is that Harrison is too impressed by the predicament of the majoritarian democrat.

Of course, I do not deny that there may sometimes be a clash between our moral commitment to the use of democratic means (which does not necessarily involve commitment to majoritarianism) and other moral commitments. And I do not deny that such clashes may have to be dealt with in the way we deal with clashes between two conflicting moral (or political) principles.

University of Edinburgh   © Vinit Haksar 1976

[1] And the relevant duty will be extinguished. See my 'Coercive Proposals', *Political Theory*, Feb. 1976, p. 76, sqq.

QUINE'S 'REAL GROUND'

By Graham Nerlich

IN RIT[1] Quine distinguishes what he calls 'the real ground' of the doctrine that translation is indeterminate from the ground provided by the *gavagai* example. He describes the real ground as follows:

> . If our physical theory can vary though all possible observations be fixed, then our translation of [a foreigner's] physical theory can vary though our translation of all possible observation reports on his part be fixed. Our translation of his observation sentences no more fixes our translation of his physical theory than our own possible observations fix our own physical theory. (RIT, p. 179–180.)

This seems straightforward. The real ground is an inference from the under-determination of scientific theory by all possible observations to the conclusion that translation is not determinate. The inference has been disputed.[2] Yet despite Quine's clear statement that this is the real ground of his doctrine, it has never attracted the same critical or elucidatory attention given to the *gavagai* example and its supporting arguments. In this note I aim at two things: first, to dispute the inference; second, to make some observations on relating this 'real ground' to other arguments in Quine.

I

The most striking fact about the inference is that it tolerates a wide divergence of views as to how much of translation it makes indeterminate. Quine is careful to point out how this flows from the fact that the real ground does not settle, by itself, how much of theoretical discourse is fixed by observation:

> What degree of indeterminacy of translation you must then recognize, granted the force of my argument, will depend on the amount of empirical slack that you are willing to acknowledge in physics. If you were one of those who saw physics as empirically underdetermined only in its highest theoretical reaches, then by the argument at hand I can claim your concurrence in the indeterminacy of translation only of highly theoretical physics. For my own part, I think the empirical slack in physics extends to ordinary traits of ordinary bodies and hence that the indeterminacy of translation likewise affects that level of discourse. But it is important, for those who would not go so far, to note the graduation of liabilities. (RIT, p. 181.)

[1] 'On Reasons for the Indeterminacy of Translation', *Journal of Philosophy* LXVII (1970), 178–183.

[2] See, for example, M. C. Bradley, 'Kirk on Indeterminacy of Translation', ANALYSIS, 36.1 (1975) 18–22; 'Quine's Arguments for the Indeterminacy Thesis', *Australasian Journal of Philosophy*, 54, May 1976, 3–31; M. Dummett, 'The Significance of Quine's Indeterminacy Thesis', *Synthese* 27 (1974), 351–397; R. Kirk: 'Underdetermination of Theory and Indeterminacy of Translation', ANALYSIS, 33.6 (1973) 195–201. I am much indebted to M. C. Bradley for valuable discussions about the present note.

Let us make use of this area of tolerance and suppose, modestly, for the sake of argument, that theory at the level of ordinary traits of ordinary bodies is determined by observation. Hence we may suppose, without offence to the inference, that language at that level of discourse translates determinately.

We take it, then, that the following sentences, at least, are determinately translatable: 'That is cubic', 'This is bouncing', 'Here is a billiard ball', 'This is like that', 'This is smaller than that' and 'I can see something spinning clockwise'. Our determinate translation of these sentences provides us, then, with a *translatable vocabulary* culled from this level of discourse about ordinary traits of ordinary things.

My challenge to the Quinean inference—to the 'real ground'—is a quite simple one. Given this vocabulary, we can write theories in it which seem very plausibly under determined by observables. To take real cases, the atomic theories of Democritus, Galileo and, to some extent, Dalton, are made up of sentences like 'The smell of cabbage is caused by a wave of cubic objects, all too small to be seen, passing up the nose' and 'A gas is really made up of lots of things like billiard balls, too small to see, hear or feel, all bouncing off one another'. Despite the high theoretical load of these sentences they are no less translatable than the sentences about ordinary objects mentioned in the last paragraph. Certainly we are speaking the language of ordinary traits of things, the traits being widely possessed by the most ordinary of objects. The objects under theoretical discussion by Democritus and Galileo are out of the ordinary only in being extraordinarily small. I can think of no plausible principle according to which what we say about them should be any less *translatable* than what we say about kitchen hardware. But, of course, what we say about them cannot be seen to be true directly. I conclude that underdetermination by observables does not imply an indeterminacy of translation.

The principle of my challenge, then, is simply that a class of sentences defined as being observably fixed may well be confined to a certain vocabulary. But it does not follow that this vocabulary is confined to the expression of sentences all of which are observably fixed.

As a last example, consider the case of two theories underdetermined by observation. Suppose a kinetic theory of gases in which a uniform direction of molecular spin (with respect to the direction of gravitational field) accounts for, say, the viscosity of gases. Theory A says that all molecules spin clockwise (seen from above) whereas theory B says that they all spin anticlockwise. No imaginable set of observations rules out either theory, let us suppose. Would there be room for doubt over which theory our Martian friends hold? We can question them about what they think in terms of standing sentences, which are written in language at a level we can translate and which includes the sentence

'This is spinning clockwise'. It is, therefore, determinate that the Martians are asserting theory A, say, and not theory B.

II

In this section I will consider some objections which might be raised against the argument just given.

(a) The Democritan sentences are not really written in the vocabulary of ordinary traits of ordinary bodies.

The crucial phrases are 'too small to be seen' and the like. We only need to unpack them a little to see that they can indeed be written in canonical notation and the terms of the level of discourse about ordinary traits of ordinary things.[1] Consider the open sentence 'For any x, I can see x iff x is larger than or equal in size to y' as a rendering of 'y is the smallest thing I can see'. (Obviously enough, spotting which *is* the smallest thing is quite another problem from translating the sentence.) Then the open sentence 'x is too small to see' is just 'x is smaller than y and y is the smallest thing I can see'. This nowhere goes beyond the vocabulary which Quine licenses us to take as determinately translatable.

(b) Quine does not say that the 'real ground' tolerates a translatable *vocabulary* at the level of discourse about ordinary traits of ordinary objects.

Admittedly, Quine does not say this in so many words, but I do not see how else he can be understood. He surely says that the real ground leaves us free to translate discourse determinately at the given level, so long as the level is defined by sentences which are fixed by observation. He cannot really mean that translation is determinate only up to stimulus synonymy. For, on the one hand, what motivates the concept of stimulus meaning is, mainly, a behaviourist approach to translation which has no obvious connection with the real ground. On the other hand, since stimulus synonymy of sentences is consistent with everything which Quine regards as really indeterminate in translation (save for inductive uncertainties) the apparent gesture of tolerance would emerge as the merest sham.

If the gesture of tolerance means anything it means at least that we can match the sentences in question one to one with foreign sentences in translation. But then we can surely parse the sentence into *terms,* unless we cannot parse it at all. It is utterly obscure how the real ground could be inconsistent with the possibility of this modest parsing, since it concerns only the relation of observation to theory and the counterpart relation of observational sentences to theoretical ones. On the face of it, it simply says nothing to the parsing issue. If parsing were ruled

---

[1] Though this requires our treating 'can see' holophrastically.

out by the real ground we would not have a situation where we *can* translate, but where it is indeterminate which translation to prefer. It would be a situation where, so far as internal sentence structure goes, we could not translate at all. This has never been what Quine has meant by indeterminate translation.

Grant, then, that the real ground tolerates a unique sentence-to-sentence translation and that translation can proceed to the matter of parsing. Then we must translate the terms of one sentence into the terms of the other. Alternative translations of terms can be in the wind only if we could match sentences equally well with alternative *sentences* containing the other terms. Now, term translation goes hand in hand with the translation of pronouns, identity, quantifiers and related apparatus. So if translation is in any real sense determinate at all it will give us a translatable vocabulary at the level in question.

(c) The argument against the inference of the real ground ignores the argumentation, surrounding the *gavagai* example, that translations term-by-term are no more determinate than translations sentence-by-sentence.

It is true that I have ignored the *gavagai* example and its immediate argumentative context. It is also irrelevant. Quine's purpose, in RIT, is just to distinguish the real ground from the *gavagai* example. But the objection above does leave us with further problems: what *is* the argumentative context of the *gavagai* example and how does it relate to the inference Quine calls the real ground?

Quine says that the real ground is 'very different, broader, deeper' (178), that it is 'another thing' (182) than the *gavagai* argument. The latter argument concerns the notion of terms and their denotation. This is 'bound up with our own grammatical analysis of the sentences of our own language' and with the problem 'what to count in the native language as analogues of our pronouns, identity, plurals and related apparatus' (181–2). It has only an 'indirect bearing on the indeterminacy of translation of sentences' (182), because a native sentence might be matched with different sentences of our own language as a function of their containing different terms (182). In short, the context of the *gavagai* example is the context of the argument from analytical hypotheses— hypotheses about the analysis of sentences (see *Word and Object*, § 15; *Ontological Relativity and Other Essays,* Ch. 1, § 1; Ch. 2, p. 31–33).

This gives us the following picture of how the real ground relates to the *gavagai* argument. Instead of Quine's metaphors of pressing from above and pressing from below we might speak instead of pressing from outside or pressing from inside the sentence. The real ground presses for indeterminacy in relating sentences to sentences in translation from outside because of underdetermination of theoretical sentences by observational ones. It is an argument wholly external to the structure

and parsing of sentences.[1] *The gavagai* argument presses for indeterminacy from inside because the indeterminacy of internal parsing and of term reference may force us to see a foreign sentence as no more clearly tied to its usual translation than to another sentence with a different term structure. That is strongly suggested, at least, by the paragraph (182) in which Quine explains why the bearing of *gavagai* on indeterminacy of sentence translation is indirect.

One remark of Quine's may cast doubt on this interpretation. Having described the scale of degrees of underdetermination and the graduation of liabilities to accept indeterminacy, he begins his discussion of *gavagai* thus: '*Gavagai,* whose troubles I shall now review, lay at an extreme of the scale' (181). This suggests that the *gavagai* argument is *a version of* the real ground, a version dealing with an extreme case on the scale. But to be on the scale at all as a version of the real ground, it would have to be seen as *a* case (an extreme case) of a sentence under-determined by observation sentences. But Quine's next remark is 'It was an observation sentence'. Any observation sentence is surely determined by an observation sentence, namely itself. I conclude that *gavagai* can-not be on the scale at all, in the sense of presenting us with a gap of the kind which might begin the inference of the real ground. Finally, I observe that the *gavagai* argument, applying as it does to observation sentences just as surely as to any others, does not permit dissent as to how much indeterminacy we must admit. It certainly cannot be under-stood to apply *only* in the higher reaches of theoretical physics.

University of Adelaide © GRAHAM NERLICH 1976

---

[1] That is another reason for seeing Quine's concession that determinacy might prevail in some areas, given the real ground, as meaning more than determinacy up to stimulus synonymy.

CORRIGENDUM

IN the article 'Meaning, Reference and Tense' by Clifford E. Williams in ANALYSIS 36.3, p. 134, the word 'not' was omitted from in front of the word 'clearly' four lines from the foot.

CAUSALITY AND OUR CONCEPTION OF MATTER

By P. J. HOLT

IN this paper I want to examine the relation between our conception of the nature of matter and the causal properties which we attribute to matter. I shall try to show that such an examination reveals certain difficulties concerning our conception of matter. I choose matter as the subject of the investigation because this seems to be the most fundamental and indispensable component of any normal person's conception of the world. It will become clear as the discussion proceeds, however, that the arguments I give could be applied with little adaptation to our conceptions of any existent things or substances other than ourselves and things modelled on our conceptions of ourselves.

I

To begin the discussion, let us consider the view—which I will call *View 1*—that the relation between our conception of matter and its causal properties is contingent. This is not a wholly implausible view. It is quite plausible to suppose that we must have notions of matter and material objects which are adequate without reference to their causal properties in order to be logically entitled to speak of such properties at all. The possession of a nature independent of its causal properties is on this view a pre-condition of a thing's entering into causal relations with other things, and there is thus no necessary relation between possession of a nature of some kind and a thing's causal properties. The theory might be regarded as an extension of the commonly accepted view that causal relations can only be discovered by empirical methods, which methods appear to demand the possession of an adequate conception of each contributing event as a pre-condition of their employment.

There are two fairly obvious objections to View 1. The first is epistemological. If things really do have "internal" or "intrinsic" natures which are quite independent of the things' causal relations to other things, then how do we—who are ourselves "other things"—come to know about these natures? On any normal theory of perception the actual perceptual experience of the perceiver is dependent on many prior causal links, and if at each stage the effect could have been caused by an agent with any nature it is difficult to see what justification we can have for trusting our supposed perceptions. Perhaps a more convincing way to put the objection, however, is this. Let us consider some of the effects commonly attributed to any familiar material object such as a chair. Two of the principal kinds of effect, which are particularly important in making the chair's presence evident to us, are those which take place on other objects trying to encroach on the chair's space and

those which act on the light striking it. Now in fact, of course, we have a notion of the space as filled with matter, and we regard this matter as responsible for the effects mentioned. But if the matter as we conceive of it does not necessarily have one effect rather than another, then why should we suppose that matter of this kind is present at all at the space in question? Why should we not simply say that there is a region of empty space through which objects are unable to pass, off which light of certain wavelengths is reflected while other light is destroyed, and so on? There is no reason to attribute these happenings to the region of space itself; they could be attributed to God, or indeed to anything or nothing. There would be grounds for .the existence of matter at the region of space only if the very nature of the matter postulated was such as to *explain* the happenings observed. But on the basis of the theory we are considering anything would do this equally well, since anything could constitute their cause.

The example of the non-causal chair can also be looked at in the following way. If our essential conception of matter does not involve causal properties, then we clearly should be able to conceive of a chair with all of these removed. It is of course logically possible, on the basis of the theory we are considering, that a chair should actually lose its causal properties while remaining unaltered in its essential nature. Suppose however that it were suggested to me that there is now a chair present before me, even though owing to its having no effects upon light I cannot see it, owing to its failure to affect material bodies I cannot feel it and can pass through it without resistance, and so on. My reaction clearly would not be that there might indeed be a chair present: rather, I should point out that what *counts* for me as a chair's being present is the ways in which, so to speak, it makes its presence felt. I would feel that there is no sense in speaking of a chair through which I can walk, which I cannot see, and so on. This plainly suggests that there is a stronger than contingent connection between our conception of matter and its causal properties.

The second objection to View 1 derives from a consideration of the properties which we commonly attribute to material objects. Let us examine some of these and consider whether any of them can shed light on our conception of the intrinsic nature of matter. One group of properties is that comprising shape, size and position. Now while it seems fairly clear that statements of position cannot describe the nature of matter, it might appear that statements of shape and size can. In fact however none of these can give us *intrinsic* properties of matter, and for similar reasons; for merely giving shape, size or position leaves completely open the nature of that which is supposed to possess the property. If I say 'Please imagine a spherical ball of radius two inches situated at the centre of London', my listener will clearly be at a loss

P 11176

what *kind* of matter to imagine; it could for example be that known as
rubber or that described as lead. My listener could even point out that
just giving him shape, size and position does not serve to distinguish a
material object from a region of empty space possessing the same pro-
perties: it is *what fills* the space that makes the difference and whose
nature we are trying to determine.

Let us therefore consider what conception we have of the difference
between, say, the matter known as rubber, that called lead, and empty
space. The essentially causal character of our notions of matter and types
of matter now emerges at once. We think of lead principally as a heavy,
solid, malleable substance, perhaps with a certain hardness and colour.
In most of these respects it is likely to differ from rubber and in all of
them it differs from empty space. Now all of these properties can be
understood only by reference to how pieces of lead move in response to
the causal influence of other things, and to how they themselves affect
other things. Strength, for example, means that the pieces of lead which
are together in a lead object do not separate easily, weight refers to the
attraction existing between matter and the earth, while colour is possessed
by an object only in that the object has the ability to affect light striking
it. It seems therefore that our very idea of what it is to be lead, or to be
matter as opposed to empty space, is essentially linked to our knowledge
of the causal effects this matter has on other things, and of the ways it
is affected by other things. This also emerged from our discussion of the
chair.

II

In the light of what has been said, it now appears that we must
reject View 1 as untenable, and adopt the view that our conception of
the nature of matter is inherently causal. According to this thesis—which
I will call *View 2*—there is a *necessary* relation between our conception
of matter and the causal properties that we attribute to matter.

Despite the apparent correspondence of View 2 with our ordinary
practice in forming conceptions of matter and types of matter this view
is itself, unfortunately, far from being obviously satisfactory in every
way. The question can be asked, for example, whether the inherently
causal conceptions of material things it postulates are logically adequate
to constitute conceptions of self-subsistent substances. An examination
of this question immediately leads to difficulties. The problem essentially
is that purely causal properties seem to tell us nothing about a substance
itself—about its intrinsic nature—at all. For causal properties surely tell
us only how *other* objects behave when—as we suppose—they are under
the influence of the thing we are trying to describe. Merely describing
something as 'that which causes x, y, z' leaves it open *what it is* that is
the cause of x, y, z. Again it appears that in order for causal properties

to get a foothold at all, we need an adequate non-causal conception at least of those things on which the effects of the causes are considered to take place. The idea that 'bundles of causes' simply act on other bundles of causes is surely logically insupportable since the notion of any cause is incomplete without the specification of its effect. Yet since material objects certainly have their principal causal effects on other material objects it seems that if our conception of the nature of matter is purely causal then our conception of the nature of the world does reduce to this idea of bundles of causes acting on other bundles of causes. We thus seem, so to speak, to lose the substance of the world; we appear on the basis of this theory to have no notion of anything which will "stand by itself"; every object seems to be dependent for what it is on other things which are themselves equally insubstantial.

Before considering View 2 in more detail and considering whether these objections can be met, I want to introduce a distinction between two kinds of necessity which could be postulated for the relation between a thing's nature and its causal properties. To explain the first, and weaker kind, let us consider the conception of a *magnet* which is likely to be possessed by an ordinary man, untrained in science. Such a man is very unlikely to know of any modification in the natures of the things he calls 'magnets' which explains their characteristic attractive properties; for him the word 'magnet' just means a piece of iron which attracts other pieces of iron. Thus although he does conceive of a magnetized piece of iron as a different sort of thing from a similar unmagnetized piece, his conception of the nature of the difference is exhausted by his conception of the difference in their causal properties. Now the kind of necessity with which magnets attract, on the basis of this conception of a magnet, might well be described as 'man-made'; it seems to arise simply from a decision to use words in certain ways. Magnets necessarily attract, on this basis, only because the definition of the word 'magnet' does not allow them not to attract.

I will describe as *View 2(a)* the thesis that there is a necessary relation of this kind between our conception of the nature of matter and its causal properties. The view implies that we have no conception of a piece of matter other than as a positioned cause. It must not be supposed that this sort of view is totally untenable. We can demonstrate its plausibility by examining that fundamental property or condition which we describe by means of the notion of electric charge. If we consider what conception we have of the difference between a charged body and an otherwise identical uncharged one, we find that this is exhaustively analysable in terms of the bodies' causal properties, of which attraction and repulsion are the principal examples. Notwithstanding our complete lack of any conception of an "internal" property in virtue of which things may be charged, however, it is undeniable

that we do normally consider charged and uncharged bodies to be genuinely different—different in themselves—even if they are otherwise indistinguishable. This is well illustrated by our conceptions of the proton and the neutron, which are certainly regarded as particles of a different nature, even though we consider that difference to reside entirely in their electrical properties. The example thus gives support to the theory that in reality all differences in actual nature which we can apprehend are causal in kind, and that any belief we might have that we can conceive of differences other than these—more genuinely "internal" differences—is illusory. This view is also supported by the following consideration. Let us suppose that a charged body does have some internal property which, as a matter of contingent fact, causes the behaviour we associate with charge. We can apparently show that this property has nothing to do with what we now understand by charge. For it would be logically possible for the body to lose the property, yet continue to display all the behaviour characteristic of charge. Since such a body would clearly be regarded as charged by us, it follows that our present conception of charge is indeed logically and necessarily linked to the causal properties of the things we describe as charged, just as the ordinary conception of a magnet is logically linked to the causal properties of magnets.

Let us now define *View 2(b)*. This is the theory that it is possible for us to form conceptions of things which, while enabling us to think of those things as self-subsistent enduring substances, are inherently causal in their natures in that the conceptions themselves render it necessary that the things should affect other things in certain ways. It could be argued that we have this kind of conception of matter itself (as opposed to particular types of matter). The above brief discussion of our conception of a chair, from which the particular importance of resistance to encroachment on a body's space emerged, suggests that the description 'space-occupier' might express our essential conception of matter. Now this description, it could be argued, embodies both existential and causal features. The existential aspect of the description is clear from the consideration that space which is occupied is occupied by something, while the causal aspect can be seen from the fact that occupancy of a region implies resistance to encroachment on that region by something else. It is certainly true that if we try to think simply of a piece of matter, we think of, so to speak, "a solid something" which not only exists at a region of space, but also by its very presence at that region implies the impossibility of the simultaneous presence of other matter. Indeed the notion 'presence at' itself seems quite an apt way to express the essence of our conception of matter; this description could be considered to have much the same existential and causal implications as that of 'space-occupier'.

We must next proceed to consider whether either or both of Views 2(a) and 2(b) can stand up to the objections brought above against any theory that there can be adequate inherently causal conceptions of matter and types of matter. We said first, it will be remembered, that an inherently causal conception does not seem to constitute an adequate conception of a substance. A substance is surely something which has its own being—a being independent of the existence of other things— and this does not appear to be true of something whose nature is essentially causal, since the very notion of a cause involves implicit reference to other things. The same kind of objection was put in purely logical terms when we suggested that there is something logically vicious in the idea of a world consisting of mere "bundles of causes".

Before treating Views 2(a) and 2(b) separately, let us consider one possible attempt to answer this logical objection to the causal theory of matter. It might be argued that our conceptions of the various constituents of the material world are logically *interdependent,* and that there is nothing "vicious" about such interdependence. In support of this claim, we could be invited to compare the constituents of the world with the elements of a mathematical system such as that of the natural numbers, 1, 2, 3, etc. It is quite possible to think of any element of this system—the number 7, say—as an entity possessing properties and thus a "nature" of its own. For example, 7 is a prime number, an odd number, and so on. But in fact of course "the nature" of the number 7 is entirely displayed by the relations it has to the other members of the system of natural numbers by virtue of its position in the system relative to them. It is a prime number, for example, because no other number except 1 divides into it exactly. Now it would clearly be a mistake to think that because our conception of each natural number is explicable only by reference to the other numbers, we are therefore involved in some kind of "vicious circle" and possess no adequate conception of any number. This reasoning is erroneous because it rests on the assumption that at some point there has to be understanding of a number *as it is in itself,* without reference to its relations to other numbers, if the logical circle is to be broken. The truth is of course that we can have understanding of the whole, interdependent system, without possessing any conception of individual members as somehow independent of it. Could it be, therefore, that we have this kind of understanding of the material world?

I think that this is quite a compelling argument, and that a realization of the apparently causal nature of all the properties (other than spatial ones) that we attribute to matter and material things almost forces us to adopt the theory it advances of our conception of the world. Nevertheless I do not believe that the argument is valid, and I think that the flaw in it lies in the inadequacy of the analogy it draws between material

things and the elements of mathematical systems. The fact is that numbers are not substances, and higher standards are required for a viable conception of the latter as existent entities than are required for such a conception of the former. Indeed the very application of the word 'existent' to numbers is inappropriate and this itself brings out the inadequacy of the analogy. We might put the matter in the following way. An adequate conception of a material object requires an understanding on our part of some *absolute difference* between an empty region of space and a region occupied by matter. This difference must lie entirely in what occurs or is present *at the region of space itself,* and it must not depend for its apprehension on any knowledge of occurrences outside that region. If this condition is not met then our conception of the world is purely formal: it consists merely in knowledge of a set of relations between entities of an unknown nature, and it is compatible with the existence of an unlimited number of qualitatively different worlds. We have just this kind of understanding, in fact, of a formal mathematical system such as a group. To a pure mathematician, a group is a system possessing structure only; it has no "content" since the nature of its elements is left entirely open. The usefulness of the concept of a group rests on the fact that many different systems of "real" entities—entities with independently understandable natures—possess its structure. Now if we really do know no more than this of the nature of the constituents of our world, if their nature, like that of the elements of a group, is entirely displayed by the relations they bear to one another, then I suggest that our supposed knowledge of the world amounts to little more than knowledge of a set of rules for predicting our perceptual experiences. And if we are in this condition there seems little point in retaining the hypothesis of an independently existing material world at all.

III

Let us now finally consider View 2(a) and View 2(b) separately. The first of these claims essentially that we have no notion of a material thing other than as a positioned cause. Thus if we are asked what it is that we are conceiving as existing at a region of space when we talk of the existence of an object at that region, we can only reply 'Something which has the following kinds of effect . . .', or in some similar way. The necessity of the causal relation here arises from the fact that we have no conception of the causing object other than as a cause; it might be argued that we have this kind of conception of that *modification* of material objects known as electric charge. In support of this view it could be pointed out first, truistically, that to think of something as a cause we are thereby necessarily thinking of it as existing, and secondly that the view does seem to be borne out by the conceptions we have of

ordinary material objects. As shown above, if we remove all the causal properties of a chair there appears to be nothing left!

Some of the unsatisfactory features of View 2(a) have already been brought out by our discussion of the logical objection to it. The difficulties attending this theory can perhaps be more forcefully demonstrated, however, by considering the following question. Are we maintaining, if we adopt View 2(a), that the existence of a body amounts to *nothing other* than the occurrence of the effects which are normally considered to be indicative of its presence, or are we claiming that some entity exists at a certain region of space which is responsible for these effects? The latter claim would of course have to amount to the postulation of some absolute difference between a region of space occupied by matter and a similar unoccupied region. Now the first alternative clearly results in the loss of the substance of the world, since the effects which are supposed to constitute the existence of a piece of matter themselves consist principally in movements of similar pieces of matter. The second alternative, however, commits us to the hypothesis of a world of individuals whose nature is to us, in our present condition, quite unknowable. It will not do to say that the natures of these entities are defined by their causal effects, for unless the things responsible for these effects have an existence of their own in addition to the effects, we are back to the first alternative once more. It seems therefore that View 2(a), while not perhaps quite untenable, certainly has consequences which must be regarded as undesirable.

Let us now turn to View 2(b). According to this, the conception we form of matter enables us to think of material things as substances existing independently of other things, but the conception is such as to render it necessary that these substances will have certain causal effects upon other things. It was suggested above that the description 'space-occupier' might express our essential conception of matter, and that this description might be considered to embody both existential and causal features.

There is certainly one obvious objection to View 2(b). This is that it seems clearly impossible to form any conception of matter which will render necessary the great majority of the causal properties we attribute to particular pieces of matter and which certainly contribute to our normal conceptions of their natures. Let us consider for example attractive properties such as electrical and magnetic properties and also that described by means of the notion of weight. There seems no possible way to conceive of a modification in the nature of an object itself—a modification in what is present at the region of space we regard as occupied by the object—which will account for these properties. Certainly there is nothing in the "space-occupier" idea which could explain attraction. There are also, however, some fairly obvious difficulties

even with the basic property of resistance to encroachment, which might seem to be clearly explained by the space-occupier conception. In the first place, this conception cannot explain *degrees* of resistance; it cannot explain the quantitative side of what occurs at a collision between objects. Secondly, we are told by scientists that there is never actual contact of matter, even at what appears to us to be a collision: what happens is that repulsive forces come into play, preventing contact and causing a rebounding effect, when bodies are very close to one another. Thus even if it is accepted that our initial conception of matter might be of the "space-occupier" kind, the question could be raised as to what should replace this, for a scientist, when he finds that such a conception does not explain the causal properties that matter actually has.

Despite this sort of objection, it might still be argued that we need not abandon View 2(b) altogether. For could we not combine it with View 2(a) to provide a compromise theory? It could perhaps be maintained that in order to get our whole conception of the world off the ground, so to speak, we need an initial conception of matter of the "space-occupier" kind. That is, a conception which logically compels a certain type, or certain types, of causal effects. The fact that matter, so conceived, has many other kinds of effect, however, would according to this theory be contingent. This would mean that when we refer to something as 'charged', 'magnetised', 'strong', 'heavy', etc., we are not referring to *intrinsic* properties of the thing—that is properties which are present at the region of space occupied by the body, and which *explain* the characteristic causal behaviour we observe; rather we are referring indirectly to the causal behaviour itself. We are thus like the scientific layman whose conception of a magnet logically compels magnets to attract iron.

That this compromise theory is untenable can be demonstrated as follows. In the first place it should be noted that it is not at all clear that we are even capable—as View 2(b) can now be seen to require—of forming a conception of an existent substance which has no properties other than occupation of a region of space. But even if it is accepted that we can do this, any causal properties of our substance which are supposed to be implied by the idea of occupation derive merely from the *logical* impossibility of the simultaneous occupation of a region of space by two numerically different pieces of matter. There are many different *practical* ways, however, in which this logical requirement could be met. For instance, two bodies, on colliding, could both be annihilated. Or both could be brought to rest. In fact, since we are concerned only with logical possibilities, we could even conceive of bizarre happenings such as the instantaneous transportation of both bodies to other unoccupied places at any distance at all from the point of collision. In the light of these considerations, it can be seen to be clear that no viable conception

can be found which renders the characteristic causal properties of matter necessary, and thus that View 2(b) must be regarded as untenable.

IV

To sum up, I will simply say that I hope I have achieved my original objective, which was to show that there are difficulties concerning the conception we can form of matter, and thus of the material world. This is not of course an original philosophical thought, but I hope that my way of trying to demonstrate that a problem does exist in this area has at least some original features.

© P. J. HOLT 1976

ON THE MERITS OF ENTRENCHMENT

By R. J. BERTOLET

NELSON Goodman claims to have given a way of solving the 'grue' problem in *Fact, Fiction, and Forecast* (Indianapolis: Bobbs-Merrill, 2nd ed. 1965; hereafter, *FFF*; parenthetical page references are to this work). I want to argue in this paper that he has not done so, by trying to show that his theory relies tacitly on the assumption that we know, roughly, that certain regularities obtain in nature, which assumption Goodman says is false (62).

My criticism departs from an important qualification which Goodman places on his entrenchment theory. Although he speaks primarily of the entrenchment of predicates, Goodman notes at one point that it is really the extension of a predicate which is entrenched. Refusing to allow any considerations which appeal to the notion of meaning to intrude upon his programme, he treats predicates purely extensionally, with the result that, for any well-entrenched predicate P and any other predicate Q coextensive with P, Q is also well entrenched. Goodman puts the point this way:

> The entrenchment of a predicate results from the actual projection not merely of that predicate alone but also of all predicates coextensive with it. In a sense, not the word itself but the class it selects is what becomes entrenched, and to speak of the entrenchment of a predicate is to speak elliptically of the entrenchment of the extension of that predicate . . . differences of tongue, use of coined abbreviations, and other variations in vocabulary do not prevent accrual of merited entrenchment. (P. 95.)

While Goodman does not explicitly say so, we can reasonably add that the new predicate is just as well entrenched as the original, by the same argument which supports its being entrenched *simpliciter*.

Besides the point that notational variation should not matter, a methodological justification for the principle is offered.

> . . . any wholesale elimination of unfamiliar predicates would result in an intolerable stultification of language. New and useful predicates like "conducts electricity" and "is radioactive" are always being introduced and must not be excluded simply because of their novelty. (P. 97.)

Refusal to acknowledge the merited entrenchment of predicates co-extensive with well-entrenched predicates would then be bad logic and bad linguistic practice, according to Goodman. This seems persuasive, and surely some form of the merited entrenchment thesis is desirable.

Now this treatment sanctions the following inference, which I shall term the merited entrenchment inference (MEI).

1. P is well entrenched.
2. $(x)(Px \equiv Qx)$

___

3. Q is well entrenched.

The question I want to discuss is: How do we know that 'green' and 'grue' are not coextensive? They had better not be, because if they are then 'grue' inherits the entrenchment of 'green' by the MEI, and 'All emeralds are grue' shares the projectibility of 'All emeralds are green'—which leaves us no way of declaring the latter but not the former law-like. And, if we are to *use* Goodman's theory to make this declaration we had better *know* that 'grue' and 'green' are not coextensive. But, this, I shall argue, is knowledge which Goodman says eludes us.

To make the discussion more precise, let us adopt a specific characterization of 'grue', applicable to time-slices of objects:

> x is grue iff x is green and t is prior to 2000 A.D., or x is blue and t is 2000 A.D. or later.

Obviously, you will say, on this understanding of 'grue' there are things in its extension which are not in the extension of 'green' (and vice versa), and so they are not coextensive predicates. Well, I think so too: but why do we think so? It might turn out to be otherwise, because the world might be different from what we think. For example, it might turn out that (1) anything which is green is green prior to 2000 and not green from 2000 on, and (2) nothing is blue from 2000 on. Suppose that this is so. Now if x is green, t must be prior to 2000, and so x is grue. If x is grue, then either x is green and t is prior to 2000, or else x is blue and t is 2000 or later. But (2) tells us that the second dis-

junct is false, and so x must be green. Thus if the world is as (1) and (2) describe it, then 'grue' and 'green' are coextensive. If 'grue' and 'green' are coextensive, then 'grue' inherits the entrenchment of 'green' by the MEI. And, if 'green' and 'grue' are equally well entrenched, we cannot discriminate between 'All emeralds are green' and 'All emeralds are grue' by Goodman's proposed criteria.

The problems which Goodman thought he had dissolved by his theory, problems about knowing what the laws or uniformities of the universe are, re-enter the picture at this point. What grounds could Goodman offer in support of the claim that the world is not as (1) and (2) specify it to be? Well, if we knew that certain uniformities obtain in nature, or that certain sentences are or express laws of nature, then we might well be able to rule this possibility out as unreasonable. But the point is that we do not know in advance of considerations of entrenchment, such as those set out above on pp. 29–30, that the uniformities we have observed will continue to obtain, or that certain sentences are or express laws of nature.

Since according to Goodman we do not know that there are these laws, etc., we do not know that the world is not describable by (1) and (2). If we do not know that the world is not so describable, then we cannot conclude that 'grue' is not coextensive with 'green'. If we cannot rule out this coextensiveness, then we cannot declare 'grue' not well entrenched, because of the MEI. Finally, if we cannot conclude that 'grue' is not well entrenched (or at least less well entrenched than 'green') we cannot rule 'All emeralds are green' projectible and 'All emeralds are grue' unprojectible. We are not entitled to say that one but not the other is confirmed.

This problem is, as stated, epistemological. If the world is described correctly by (1) and (2), then 'All emeralds are grue' is confirmed by Goodman's criteria—which he would say is just as it should be. If the world is as we think it is, the 'grue' hypothesis is ruled out; the problem is that if we simply accept Hume's claims, as Goodman does, we give up the possibility of knowing that it is ruled out. So the problem is not that Goodman's theory gives the wrong results, but that we cannot know what results it provides. The 'grue' hypothesis may or may not be confirmed, but by Goodman's theory we cannot tell which. My claim, then, is simply that Goodman's theory does not give us a way to say that 'All emeralds are green' is confirmed by the inspection of green emeralds while 'All emeralds are grue' is not; yet the theory was designed to enable us to do just that.

University of Wisconsin—Madison   © R. J. BERTOLET 1976

THE FOUR-DIMENSIONAL WORLD

By H. W. NOONAN

MANY eminent philosophers—examples are Nelson Goodman, David Lewis, Richard Montague, Quine and J. J. C Smart—take the view that enduring objects can be thought of as 'four-dimensional worms' or 'timelike streaks'.[1] But there is an argument of Geach's against this view. The primary purpose of this paper is to offer the four-dimensionalist an answer to this argument. Here is the argument:

> . . . another example: 'McTaggart in 1901 was a philosopher holding Hegel's dialectic to be valid, and McTaggart in 1921 was a philosopher not holding Hegel's dialectic to be valid'. If we regarded 'McTaggart in 1901' and 'McTaggart in 1921' as designating two individuals, then we must also say they designate two philosophers: one philosopher believing Hegel's dialectic to be valid, and another philosopher believing Hegel's dialectic not to be valid. To be sure, on the view I am criticizing, the phrases 'McTaggart in 1901' and 'McTaggart in 1921' would not designate two philosophers, but two temporal slices of one philosopher. But just that is the trouble: for a predicate like 'philosopher believing so-and-so' can of course be true only of a philosopher, not of a temporal slice of a philosopher. So if our example, which is a plain and true empirical proposition, were construed as a conjunction of two predications about time slices of McTaggart, then it would turn out necessarily false; which is an absurd result. The absurdity does not come about just for my chosen example; it arises equally for Quine's example 'Tabby at *t* is eating mice'; for a cat can eat mice at time *t*, but a temporal slice of a cat, Tabby-at-*t*, cannot eat mice anyhow.
>
> The friends of temporal slices will no doubt here pray leave to amend the examples so that they contain predicates fitting temporal slices, instead of predicates like 'philosopher believing so-and-so' or 'cat eating mice', which fit living beings. . . . But we ought not to grant them leave to amend. The whole ground for treating, for example, 'McTaggart in 1901' and 'McTaggart in 1921' as designating two distinct individuals was that we seemed to find predicates true of the one and false of the other. But now we find that such predicates as appear in ordinary empirical propositions are often of a kind that could not be true of temporal slices; so the ground for recognizing temporal slices as distinct individuals has been undercut: and we ought to reject temporal slices from our ontology, rather than cast around for new fashioned predicates to distinguish them by.[2]

My answer to this argument is that McTaggart-in-1901 and McTaggart-in-1921 *are* both of them philosophers, and that, despite the fact that different predicates are true of them, they are the same philosopher.

[1] Lewis' expression—see his recent paper, 'The Paradoxes of Time Travel', *American Philosophical Quarterly*, Vol. 13, No. 2, April 1976.
[2] *Logic Matters*, Oxford: Basil Blackwell, 1972, p. 310.

This is not the only reply to Geach's argument which is available to the four-dimensionalist. J. J. C. Smart has suggested another.[1] Essentially his answer amounts to denying that from the fact that something believes in Hegel's dialectic one can infer that it is a philosopher, for it may merely be a temporal slice of a philosopher. He thus holds that 'McTaggart in 1901 believed in Hegel's dialectic' *may* be analysed as attaching the predicate 'believes in Hegel's dialectic' to the time-slice designation 'McTaggart-in-1901', but that 'McTaggart in 1901 was a philosopher believing in Hegel's dialectic' cannot be similarly analysed and must instead be paraphrased by something containing predicates which *can* be sensibly attached to time-slice designations (he gives reasons for rejecting Geach's claim that such paraphrasing is illegitimate).

As an answer to Geach's argument Smart's suggestion seems perfectly adequate. But I have two reasons for preferring my own. First of all, it also serves as an answer to an objection of D. H. Mellor's to the four-dimensionalist position, which objection, I think, it is very difficult, if not impossible, to answer if one adopts Smart's suggestion. And secondly, answering Geach in my way makes it possible for the four-dimensionalist to give an account of the distinction between process terms (like 'wedding', 'lecture', 'war', 'noise') and continuant terms (like 'man', 'table', 'omelette', 'car'). We all perceive a distinction here, and philosophers like Geach and Mellor are ready with an account of it: according to them it is the distinction between those terms which signify things with temporal parts, and those terms which signify things without temporal parts. The four-dimensionalist cannot accept this account of the distinction, of course, but nor can he expect his opponents to abandon it unless he offers them something in its place.

First, then, I shall elaborate on and defend my answer to Geach, then show how it is also an answer to Mellor's objection, and finally explain my account of the distinction between process terms and continuant terms.

My answer to Geach, to repeat, is the following: I reject his claim that the predicate 'is a philosopher' cannot be true of the time-slices McTaggart-in-1901 and McTaggart-in-1921, and hold that in fact *every* time-slice of McTaggart during his philosophical career satisfied that predicate. And, consequently, to avoid absurdity, I also hold that McTaggart-in-1901 *is the same philosopher as* McTaggart-in-1921 despite the fact that they have different properties. To say this, of course, is to take a leaf out of Geach's own book[2] for it involves holding that *being the same philosopher as* does not entail Leibnizian identity (i.e. the relation which satisfies Leibniz's Law) so that 'is the same philosopher as' cannot

[1] J. J. C. Smart, 'Space-Time and Individuals', in Rudner and Sheffler edd., *Logic and Art*, New York: Bobbs-Merrill, 1972, pp. 12–13.
[2] *Reference and Generality*, Ithaca: Cornell University Press, 1962, p. 151.

be thought of as split up into 'is a philosopher' and 'is the same as'. And sustaining this position must, I think, involve still further borrowings from Geach—since one cannot regard 'is the same philosopher as' as logically complex, composed of a relational expression 'is the same . . . as' and the general term 'philosopher', one must regard it as logically simple, like 'is a brother of', and regard 'is a philosopher' (or better 'is a man') as a derelativization of 'is the same philosopher (man) as' just as 'is a brother' is a derelativization of 'is a brother of'.[1] But I see no reason why these latter borrowings should be unacceptable to a four-dimensionalist, especially in view of Quine's recent acceptance of them.[2]

But a problem now arises: if McTaggart-in-1901 and McTaggart-in-1921 are one and the same philosopher can we not speak of the (one and only) philosopher who is the same man as McTaggart-in-1901 and is the same man as McTaggart-in-1921? Yet, if we can, must it not be true both (1) that the one and only philosopher who is the same man as McTaggart-in-1901 and is the same man as McTaggart-in-1921 is a believer in Hegel's dialectic (since McTaggart-in-1901 is) and (2) that the one and only philosopher who is the same man as McTaggart-in-1901 and is the same man as McTaggart-in-1921 is not a believer in Hegel's dialectic (since McTaggart-in-1921 is not) without its being true (3) that the one and only philosopher who is the same man as McTaggart-in-1901 and is the same man as McTaggart-in-1921 both is a believer in Hegel's dialectic and is not a believer in Hegel's dialectic (for neither McTaggart-in-1901 nor McTaggart-in-1921 nor any other temporal slice of McTaggart satisfied the conjunctive predicate 'both is a believer in Hegel's dialectic and is not a believer in Hegel's dialectic')? But is this not impossible? For 'the (one and only) philosopher who is the same man as McTaggart-in-1901 and is the same man as McTaggart-in-1921' is a definite description (it is the four-dimensionalist equivalent of 'the philosopher who was the same man as McTaggart was in 1901 and the same man as McTaggart was in 1921'), but definite descriptions satisfy the *conjunction* test: that is, if the A which is F is G and the A which is F is H, the A which is F is (G and H).

I can only think of one way of dealing with this problem while maintaining that McTaggart-in-1901 and McTaggart-in-1921 are each of them philosophers, but the *same* philosopher, and this involves yet further borrowings from Geach.

What the four-dimensionalist must do, I think, to deal with this problem is to reject the customary view that statements beginning 'some man is . . .' and 'every man is . . .' are paraphrasable, respectively, as 'something is a man and is . . .' and 'everything, if it is a man, is . . .'[3]

---

[1] See Geach's paper, 'Ontological Relativity and Relative Identity', in Milton Munitz, ed., *Ontology*, New York: New York University Press, 1973.

[2] See *The Roots of Reference*, La Salle, Illinois: Open Court, 1973, p. 57.

[3] cf. *Reference and Generality*, pp. 150–151.

He must, in other words, claim that the restriction conveyed by 'some man' and 'every man' is not just a restriction to those things which satisfy the predicate 'is a man' (for every temporal slice of a man satisfies that predicate). But what, then, *is* it a restriction to? The plausible answer to this is fairly obvious: it is a restriction to those things which are men *and are not proper temporal parts of anything which is a man*—it is a restriction to McTaggart and not just to McTaggart-in-1901 and Mc-Taggart-in-1921. And, clearly, if this were not so then many of the things we believe to be true would be false: it is true, I suppose, that every man has existed for more than five minutes (forgetting pre-natal fatalities), but it is not true, even forgetting pre-natal fatalities, that every person-stage has existed for more than five minutes—far from it.

The restriction in question can be stated in two other ways. We might say: person-stages which are the same man are related by a relation of "personal unity". Then the restriction is a restriction to those summations of person-stages related pairwise by personal unity *which are not proper parts of any other* such summations. Or the restriction can be stated in this way: it is a restriction to the things x satisfying the following condition: x is a man and for any y, for any t, for any t', if x at t is the same man as y at t', x at t' is the same man as y at t'.

It may help here if I make the following comparison. If one thinks of colours in Quine's way, so that one regards 'Red' for example, as naming all the red stuff there is, i.e., the spatio-temporally scattered totality of red substance, then Red, like any of its temporal parts, will *be red* and so will not only be a colour, but will also *have* a colour. It will be distinguished from any temporal part of it, however, in that it will not be a part of anything which has a colour, or equivalently by the fact that it, but none of its temporal parts, will satisfy this condition: x has a colour, and for all y, for all t, for all t', if x at t has the same colour as y at t', x at t' has the same colour as y at t'. The same holds for the other colours, construed in Quine's way. The restriction conveyed by 'some colour' and 'every colour' may thus on this view be regarded equally as a restriction to those things satisfying the predicate 'x is a colour' or to those things which have colours, *but are not proper parts of anything which has a colour*. Similar remarks hold for Gold, construed in Quine's way. Thus construed it is the spatio-temporally scattered totality of gold substance, and so *is gold* in just the same sense as any temporal part of it (forgetting those parts which are too small to count). Consequently it *is a metal* in just the same sense as they are (whereas in the case of colours we speak of samples of a colour as *having* a colour rather than as *being* a colour, we unfortunately make no such distinction in the case of metals—to speak of something as 'having a metal' is not English). It is distinguished from any temporal part of it, however, by the fact that it, but none of its temporal parts, is not a part of anything which is a

metal, or equivalently by the fact that it, but none of its temporal parts, satisfies this condition: x is a metal, and for all y, for all t, for all t', if x at t is the same metal as y at t', x at t' is the same metal as y at t'. Similarly for the other metals, construed in Quine's way. The restriction conveyed by 'some metal' and 'every metal' may thus on this view be regarded as a restriction to those things which are metals, *but are not proper parts of anything which is a metal*. My suggestion is that the relation of 'some man' and 'every man' to 'is a man' is parallel to the relation here between 'some colour' and 'every colour' and 'has a colour' or 'some metal' and 'every metal' and 'is a metal'.

Now clearly, any statement about 'the philosopher who . . .' is in the first place equivalent to something of the form 'some philosopher is . . . and every philosopher who is the same man as he is . . . and - - -' which uses restricted quantification, and an analysis of it along Russellian lines using 'something is a philosopher and . . .' and 'everything, if it is a philosopher . . .' will be adequate only if it accords with that equivalence. But, if the restriction conveyed by 'some man' and 'every man' is the one we are at present assuming, it cannot do so. So the solution to our problem is this: we can indeed speak of 'the (one and only) philosopher who is the same man as McTaggart-in-1901 and is the same man as McTaggart-in-1921', but since any statement containing this description will be equivalent to something involving the restricted quantifiers 'some man' and 'every man' it is satisfied neither by McTaggart-in-1901 nor by McTaggart-in-1921 nor by any other temporal slice of McTaggart, but only by *McTaggart*. Now it is true of McTaggart that in 1901 he believed in Hegel's dialectic (since McTaggart-in-1901 believed in Hegel's dialectic), and true of McTaggart that in 1921 he did not believe in Hegel's dialectic (since McTaggart-in-1921 did not believe in Hegel's dialectic), and it is also true of McTaggart, of course, both that in 1901 he believed in Hegel's dialectic and that in 1921 he did not. The description 'the (one and only) philosopher who is the same man as McTaggart-in-1901 and is the same man as McTaggart-in-1921' does therefore pass the conjunction test after all.

Now a few more words about the suggestion that the restriction conveyed by 'some man' and 'every man' is not merely a restriction to those things which are men, i.e. satisfy the predicate 'is a man'. This may seem a strange suggestion, but there is nothing incoherent about it— it is just an empirical hypothesis about English, and, I hope to have shown, one that must be accepted if answering Geach in the way I suggest is to be regarded as acceptable. It should be carefully distinguished, however, from Geach's own suggestion on pp. 150–151 of *Reference and Generality*. The difference is this: while the present suggestion is that the restriction conveyed by 'some man' and 'every man' cannot be regarded as a restriction to those things satisfying the predicate 'is a

man', it does not imply that the restriction cannot be regarded *at all* as a restriction to those things satisfying a certain predicate—on the contrary, it is part of the suggestion that the restriction is to those things satisfying the predicate 'is a summation of person-stages linked pairwise by personal unity which is not a proper part of any other such summation'. By contrast, it is an essential part of Geach's view that *no* predicate could capture the restriction conveyed by 'some river' or 'some water'; in short, his view is that such restricted quantification is *absolutely* irreducible to unrestricted quantification, and his unFregean advocacy of *general names* as a category distinct from predicables is connected with this. I am not suggesting that the four-dimensionalist should depart as far as Geach does from the usual view of the reducibility of restricted quantification to unrestricted quantification, but only that he should occupy a midway position.

I now turn to Mellor's objection:

> If a thing *is* [a causal sequence of events] it has temporal parts ('phases'), namely the events in the sequence; the ontological distinction between things (substances) and events is destroyed, since it is just lack of temporal parts that distinguishes a thing from temporally extended events. Well, why not destroy the distinction—perhaps there are no things without temporal parts? But if so, no temporally extended object is wholly present at a time; conversely, what is present at a time is not wholly identical (and so, perhaps, *pace* Parfitt, not identical at all) with anything at any other time. And where then, for example, in the case of people, do our notions of moral and of legal responsibility for what *we* have *ourselves* done, and intended ourselves *to* do, go?[1]

Certainly, in the context, say, of a trial, a crucial question is whether the man in the dock is the same man as the man who, say, robbed the High Street Jewellers six months ago; but the four-dimensionalist can accommodate this quite easily, at least if he answers Geach in the way I have suggested. And evidently, once it has been established that the man in the dock is the same man as the one who committed the robbery, no further investigation is considered necessary to determine whether he is *identical* with him.

Mellor's objection would have force, I think, if, as the first part of Geach's argument assumes, individuals with distinct properties could not be the same man, for then to punish Jones-at-*t* for what was done by Jones-at-*t'* would in all cases be to punish a thing for what was done

---

[1] D. H. Mellor, review of J. L. Mackie *The Cement of the Universe: A Study of Causation*, *Ratio*, Dec. 1975. In an unpublished paper called 'Social Individuals' Mellor makes some remarks which make the point of this argument a little clearer. He argues that, because we ascribe legal and moral responsibility to them, such things as companies, which he calls 'social individuals', should be regarded as continuants and not processes. He writes 'A social individual, like a person, is responsible only for what *it* does or leaves undone. Here strong self-identity over time is even more plainly necessary. The Distillers Co in 1973 could not be responsible for The Distillers Co in 1960 being slow to withdraw thalidomide were it not the same company; in just the same sense in which only the same person can be responsible for his past actions.'

by something which was *not the same man as* it (Jones-at-t and Jones-at-t' would either be different men, or not men at all), and that does seem to go against our conceptions of responsibility, both moral and legal. But this just means that Mellor's objection provides yet another reason for the four-dimensionalist to reject that assumption, and hold to the position elaborated in this paper, that every time-slice of a man is a man and the same man as every other time-slice of him.

Finally I turn to the distinction between process terms like 'wedding', 'lecture', 'war', 'noise', and continuant terms like 'man', 'table', 'omelette'. What is the distinction we all perceive here? Well, note that all these words can occur in three contexts: all of them can take the place of 'A' in 'some A', 'every A' (and also 'most A', etc.,), all of them can take the place of 'A' in 'is the same A as', and all of them can take the place of 'A' in 'x is an A'. My suggestion then is that 'A' is a continuant term if and only if (1) every temporal part of something which is an A is itself an A, (2) every temporal part of something which is an A is the same A as it, and as every other temporal part of it, and (3) the restriction conveyed by 'some A' and 'every A' is a restriction not merely to those things which satisfy the predicate 'is an A', but, more narrowly, to those things which are A's *and are not temporal parts of anything which is an A.*

As we saw in discussing Geach's argument, (3) is not independent of (1) and (2): if 'A' satisfies (1) and (2) then it could only fail to satisfy (3) if definite descriptions of the form 'the A which . . .' failed the conjunction test, and it is evident that no definite descriptions, either of continuants or of processes, do fail this test. It is clear then that condition (2) is the crucial one: although some process terms satisfy condition (1) ('noise' is perhaps an example), none of them satisfies condition (2).

To explain the distinction between process terms and continuant terms in this way evidently concedes nothing to Mellor or Geach, but it accommodates Mellor's intuition, which is widely shared, that continuants are self-identical over time in some strong sense in which processes are not, and seems to fit in quite well with the way we actually talk. It is natural to use temporal qualifiers with terms we intuitively regard as signifying continuants, but not—with one important class of exceptions—with terms we intuitively regard as process terms: 'is a fat man now' and 'was a fat man last year' seem perfectly natural expressions;[1] 'is an interesting lecture now' and 'was an interesting lecture half an hour ago' seem much less natural. The class of exceptions are process-terms like 'noise', which seems to satisfy condition (1), and this,

---

[1] The point here becomes still clearer when one thinks specifically of attributive adjectives : it is perfectly natural to say things like 'John was a good man before he met Bill (started to drink, won the pools)', and here, because of the well-known facts about attributive adjectives, the use of the substantive 'man' in association with the temporal qualifier is indispensable.

of course, is just what one would expect. The reason for holding that continuant-terms satisfy condition (1) is just that this provides the simplest explanation compatible with the four-dimensionalist view-point of the linguistic facts just referred to. The facts about the way we count continuants over time then force one to recognize that they also satisfy condition (2), and the fact that definite descriptions pass the conjunction test leaves one with no choice but to accept that they also satisfy condition (3). It is, of course, just because we do so very often say things of the form 'N was an F man at t but a G man at t'' that Geach's argument is able to get started at all.

Trinity Hall, Cambridge   © H. W. NOONAN 1976

CHANDLER ON THE CONTINGENTLY POSSIBLE

By R. A. FUMERTON

IN 'Plantinga and the Contingently Possible' (ANALYSIS, 36.2) Hugh S. Chandler argues that there are states of affairs which are only contingently possible (possible in some possible worlds and not in other). He relies, in effect, on the intuitive plausibility of a specific example. He claims that some physical object (he calls it 'Alfred') made up of parts E_1, E_2 and E_3 might be such that it could have come into existence with one of its three parts different even though it could not have come into existence with two of its three parts different. If one can make the claim plausible by appealing to what we would say of everyday objects one would have made plausible the claim that there are states of affairs which are only contingently possible, for in our world it is true that Alfred might have come into existence with parts E_1, E_2 and E_4 but not true that Alfred might have come into existence with parts E_1, E_4 and E_5. This is only a contingent fact, however, for if Alfred had come into existence with parts E_1, E_2 and E_4, it would have been possible for it to have come into existence with parts E_1, E_4 and E_5.

Is it true that it is either *possible* for something (say my bicycle) to have come into existence with one of its parts different or *impossible* for something to have come into existence with all of its parts different? Consider for a moment identity through time. If we say that the bicycle I own at t_1 made up of three parts E_1, E_2 and E_3 would remain that same bicycle at some later time only if it has at that later time at least two of the three parts it had at t_1, then we have committed ourselves to an absurdity —transitivity of identity fails through time. If at t_1 I have a bicycle made

up of E_1, E_2 and E_3, at t_2 I replace E_3 with E_4, and at t_3, replace E_2 with E_5, then it will be (1) true that the bicycle I own at t_1 is identical with the bicycle I own at t_2, and (2) true that the bicycle I own at t_2 is identical with the bicycle I own at t_3, but (3) false that the bicycle I own at t_1 is identical with the bicycle I own at t_3. Chandler must either admit that there is a serious disanalogy between questions concerning identity through possible worlds and questions concerning identity through time or give up the crucial premiss of his argument. He clearly chooses the first alternative—'A bicycle can survive the gradual replacement of each and every one of its parts'.[1] But now how are we to test our intuitions with respect to Chandler's claims about the identity of certain things through possible worlds? We certainly do make claims about the identity of things through time, but such intuitions are useless for resolving our present problem; for as noted above the criteria for identifying things through time must be different from the criteria for identifying things through possible worlds. Could that *very same* bicycle which is at a time *t* in this world made up of parts E_1, E_2 and E_3 have been made up at that time in some other world of parts E_1, E_2 and E_4? Is it impossible for that very bicycle made up of parts E_1, E_2 and E_3 in this world to be made up of completely different parts in some other world? I do not believe there are any coherent criteria for determining the answer to such questions. I suppose it seems odd to suggest that that same bicycle could have come into existence with *all* of its parts different, but on the other hand it seems at least initially odd to suggest that the bicycle I own now can be identical with the bicycle I own at some later time when *all* of its parts are different. In identifying the thing through time the *gradual* character of the change obviously strikes us as making a difference and I suspect that what makes identity judgments with respect to things through possible worlds so difficult is precisely the fact that there is nothing analogous to a change taking place *gradually* (where this is clearly a temporal relation). Though it is too controversial a conclusion to fully argue here, I suspect that we get along as well as we do talking about the identification of things through possible worlds, because we view it as analogous to the identification of things through time. When the disanalogy is revealed I hope that philosophers will begin to see more clearly the futility in continuing to act as if there were any coherent criteria for the identification of things through possible worlds.

University of Iowa   © R. A. FUMERTON 1976

TENSES AND MEANING CHANGE

By Stephen E. Braude

IN 'Meaning, Reference and Tense' (ANALYSIS, 36.3), Clifford E. Williams criticizes a position I advanced but did not defend in a paper that appeared a few years ago.[1] In that paper I considered and rejected various ways of distinguishing tensed from tenseless sentences. One of these ways is to maintain that nonsimultaneous replicas of a tensed sentence differ in sense (or express different propositions), whereas nonsimultaneous replicas of tenseless sentences do not. It seemed to me then, and still does, that this position is patently false. I provided no argument against the position, however, and instead merely asserted that it is a "brute fact" that successive replicas of such tensed sentences as 'J.F.K. was assassinated' can have the same sense.

Williams charges correctly that I offer no defence of my claim that this is a brute fact about language. But I believe that Williams has failed to see just what sort of fact this brute fact is. I do not intend to offer a fully fledged defence of my claim that my alleged brute fact is a fact, since to do so would require a lengthy account of the nature and development of natural languages generally. But I believe I can make clearer what sort of claim I am making about language.

Williams asserts that I never explain *via* any theory of meaning what it is for two sentences to have the same sense. Unlike Williams, however, I see this as a virtue of my approach. I think it is a mistake to begin an enquiry into tenses with some philosophical theory of what sentences express, and then try to decide whether tensed sentences can express the same thing at different times. A theory of meaning must stand or fall, it seems to me, on the basis of its compatibility with certain pretheoretic assumptions about language use. One of these assumptions is that people can express what they themselves or others have previously expressed by replicating the sentences used at those earlier times. Most of the sentences used in our everyday communicating are tensed, and with respect just to tensed sentences this assumption amounts to saying that replicas of such ordinary sentences as 'Cincinnati won the 1975 World Series' and 'Mary is sad' can express the same thing at different times, when produced either by the same person or by different persons. This is simply part of the data upon which any satisfactory theory of meaning applicable to natural languages must be based. Any language for which this assumption fails is simply not a natural language.

Remember that it takes hardly any linguistic competence at all for someone to express what he or someone else previously expressed. A

---

[1] 'Tensed Sentences and Free Repeatability', *Philosophical Review*, Vol. 82, No. 2 (April 1973).

child can do it. Suppose that a young boy points to himself and says 'Hungry!', and then, seeing that his parents did not hear him, points to himself again and says 'Hungry!'. Had the child been more expert in speaking the language he might have said 'I am hungry'. But had he produced the more complicated string of sounds he would have expressed nothing more nor less than what he expressed with 'Hungry!'. Now it seems that we must suppose that the child expressed exactly the same thing with his two rudimentary tensed sentences if we are to be talking about a natural language. Any language which failed to incorporate such a simple device for expressing what has been previously expressed would clearly not meet the urgent needs for communication that motivate the development of a natural language in the first place, and it probably would not be a language which idiots and children could use.

I must also protest against Williams' use of a sentence like

(1) The man in black is reading a book now

as a paradigm tensed sentence. Paradigm present-tense sentences do not contain the demonstrative 'now'. Philosophers usually insert the word 'now' into ordinary present-tense sentences to make explicit what they already regard as a reference to the present contained in the sentence. But, since there are those who (like myself) think that in general tensed sentences do not refer to times, this procedure is question-begging. A paradigm tensed sentence is more like

(2) Jones is sick

(3) Smith was born in Boston

and it is not at all obvious that these sentences must express different things at different times.

But even (1), it seems, can have the same sense at different times. Imagine the following dialogue.

Speaker A: Look! The man in black is reading a book now. Thank God he's stopped trying to molest the maid.

Speaker B: Well, let me see. Yes, you're right. The man in black is reading a book now.

Furthermore, whatever precisely is meant by saying that the replicas of (1) in this dialogue express the same thing, it is clearly not that they have the same rules of use. Williams suggests I may be presupposing such a view. But that the two replicas of (1) express the same thing has to do with such facts as that 'The man in black' refers to the same person in both cases. And it is not part of the rules of use of the expressions used by Speakers A and B that anyone in particular gets referred to by 'The man in black'.

It is curious, incidentally, that after pointing out that I do not explain what the *sense* of a sentence is Williams goes on to talk about the *content* of a sentence without explaining what *that* is. How does it differ from a sentence's sense? In terms of what theory of meaning are we to understand contents?

Finally, the closest thing Williams offers to an argument against the acceptance of my "brute fact" is this. He writes,

> . . . All [indexical sentences] can have different truth-values when produced in different circumstances, and in order to know what their truth-values are, we have to know what circumstances they are produced in. So, it would seem that (1), too, expresses different contents and refers to different states of affairs at different times. (P. 133.)

As I understand it, the argument, in a more general form, is this.

(i) Replicas of indexical sentences can take different truth-values in different circumstances of production.

(ii) To know the truth-value of an indexical sentence, we must know its circumstances of production.

(iii) *Therefore,* indexical sentences express different contents and refer to different states of affairs in different circumstances of production.

First of all, it is not clear what Williams has in mind when he writes of sentences *referring* to states of affairs. Most tensed sentences do not refer (or even contain references) to states of affairs, though probably all such sentences *report* states of affairs. But, if Williams is concerned with the reporting of states of affairs, then nonsimultaneous tensed sentences can clearly report the same thing and (iii) is false. Consider, e.g., successive replicas of (3): both sentences report Smith's birth. Granted, the different truth conditions of the two sentences may relate Smith's birth to different times (i.e., the times of production of the sentences). But the two sentences do not refer, nor do their truth conditions obviously refer (or perhaps better, contain references), to different states of affairs. Perhaps Williams is saying that a change in a sentence's truth conditions requires a change in proposition expressed. This is a familar view about propositions, and one which I clearly deny. But the above argument does not establish this view.

If we focus on the other half of the conclusion of Williams' argument—namely, that indexical sentences express different contents at different times, then the argument is simply a *non sequitur.* We can see what is wrong by observing that not all indexical sentences are tensed. Consider the *tenseless* schema

(4) Event E occurs [tenseless] here.

Replicas of instances of (4) *can* have different truth-values, when produced at different places. But they express different contents only in different places. So, although replicas of an instance of (4) can have different truth-values in different contexts, and although we must know what the context is to know what a replica's truth-value is, it does not follow that in different contexts replicas have different contents. Williams thus fails to establish any interesting link between the need to know a sentence's context of production to know what the sentence's truth-value is and what it is that the sentence expresses in that context.

Moreover, there are, as I noted in my article, tensed sentences which *cannot* change in truth-value with time—e.g.,

(5) Nobody has been a married bachelor

(6) I do not exist.

If Williams wants to claim that nonsimultaneous tensed sentences do not express the same proposition, or content, or whatever, he must do so on some basis other than the possibility of a truth-value change between replicas of tensed sentences, since these sentences are necessarily true and false, respectively.

University of Maryland Baltimore County   © STEPHEN E. BRAUDE 1976

ASSERTION: A REPLY TO BROOKS

By MICHAEL COHEN

ACCORDING to D. H. M. Brooks, an account of a language in terms of the sense and reference of its sentences merely distinguishes sentences into two classes 'without specifying which of these two classes the language speakers attempt to utter' (ANALYSIS, 36.3, p. 117). We can imagine that a linguist gives us the following information about the language of a tribe: that it contains a sentence $\ulcorner Fa \urcorner$ which belongs to a certain class V of sentences if and only if the object a is round. The linguist however fails to tell us whether V is the class of sentences which these people regard it as correct to utter or the class they try to avoid uttering. Brooks is under the impression that I deny we could have such knowledge of a language. But I find no difficulty in conceiving this. What I deny is Brooks's claim that this knowledge amounts to knowledge of the sense of $\ulcorner Fa \urcorner$.

One obvious objection to Brooks's idea of sense is this. Imagine the language has a sentential connective '*' whose behaviour is described thus: $\ulcorner P * Q \urcorner \in V$ if and only if $P \in V$ and $Q \in V$. Since we do not know which class V is, we do not know whether '*' is a sign for con-

junction or for disjunction; but we have determined what Brooks would call the *sense* of the sign.

As Brooks points out, it may be that there are two tribes who use exactly the same sentences, but in opposite ways, one tribe taking V as the class of sentences to be uttered, the other taking V as the class to be avoided. In the one language, then, he tells us, 'F' means 'round' while in the other it means 'not round'. But 'F' has the same *sense* in both languages.

The difficulty here is to see what role this notion of sense is supposed to play in an account of language. Brooks tells us at one point that the extension of the predicate 'F' in one language is the complement of its extension in the other. But if *extension* though not *sense* is relative to a language we are at a loss to explain the relation between them. We can no longer say that to know the sense of a predicate is to know how to determine its extension, for there is no such thing as the extension of a predicate—only its extension relative to a language. An analogous point can be made with respect to the sense of a sentence and its truth-value; if truth and falsity are relative to a language while sense is not, sense cannot be identified with truth-conditions. (In fact, for reasons which I do not properly understand, Brooks wishes to avoid the notion of truth. But the argument goes through if we substitute his notion of the correctness of an utterance.)

At the end of his paper, Brooks remarks that the main reason for holding that an account of a language in terms of the sense and reference of its sentences is inadequate is that 'it enables us to deal with assertion on a par with other types of linguistic force, such as imperative or interrogative force' (*loc. cit.*). But from the fact that one has got hold of something which one calls the sense of the sentences of a language, which one can be said to grasp without knowing what it is to make assertions in the language, it doesn't follow that 'sense' represents something common to both assertions and commands. Indeed Brooks's own description of sense fails to meet this demand, for while we can see what needs to be added to what he calls a sense-reference account for us to know what it is to assert something in the language (that is, a specification of which class of sentences the speakers of the language try to utter) there is no parallel for commands. In any case Brooks gives us no reason for supposing that there is something common to the command, say, to shut the door and the assertion that the door is shut, beyond the rather trivial fact that one could not be said to understand the command unless one also understood the assertion. He is of course assuming a view common in recent philosophy. It would take another paper to say what is wrong with this view.

University College of North Wales, Bangor   © MICHAEL COHEN 1976

WOODRUFF ON DISCRIMINATION

By STANLEY S. KLEINBERG

COMPENSATORY discrimination in favour of a group will often benefit individuals who have not themselves been wronged. Is it not therefore unjustified? Paul Woodruff (ANALYSIS, 36.3, pp. 158–160) argues that it is not and that compensation is owed equally to all the members of any group whose respect is diminished by the existence of a pattern of discrimination. I contend that his account leads to serious difficulties, and that a better understanding of the nature of discrimination is needed if we wish to cast light on compensatory discrimination.

The central feature of Woodruff's account is his claim that what is wrong with discrimination (when it is wrong) is to be found, not in the fact that individuals are unfairly denied opportunities, but in the insult to a group (or unfair damage to the respect in which the group is held) which occurs only where a general pattern of discrimination exists. I shall refer to this as the Group Insult Claim (GIC).

We might note in passing that Woodruff's espousal of the GIC, together with his conclusion on compensatory discrimination, raises the question of whether he should not be advocating compensation for groups who are merely subject to prejudice. This is because behaviour that is insulting to groups is clearly not manifested exclusively in overt acts of discrimination. If whites move out of an area because they do not wish to have black neighbours or vote for racialist politicians, this is insulting. If the GIC is right, then such happenings are morally equivalent to patterns of discrimination, and we might expect Woodruff to hold that compensation is due.

Is the GIC internally consistent? A supporter is committed to holding that isolated acts of discrimination are not wrong. This is plausible if we think that what is wrong with discrimination resides in its unfair denial of opportunities to individuals. In that case if discrimination is practised by one bank amongst many and there are ample vacancies in banking, then would-be bankers in the group that is discriminated against suffer no significant loss of opportunity. But the GIC asserts that patterns of discrimination are wrong because they are insulting. In order to avoid self-contradiction the GIC must be construed as holding that particular types of behaviour can be saved from being insulting by virtue of not being part of a pattern. This is indeed Woodruff's view; moreover he makes the question of whether an act is insulting turn on whether or not it is the reason for failure of the person who is its 'victim'. He writes as follows:

> If an applicant fails at one bank because of his race, and at other banks for other reasons, his race is not the reason for his unemployment, and his failure is not an insult to his race. (p. 159.)

46

My view is that his race is both *the* reason for his failure at one bank and *a* reason for his unemployment (depending on the circumstances it could be the most important reason), and that whether the discrimination is an insult does not turn on whether the applicant finds alternative employment. If my understanding of the normal meaning of 'insult' is correct, some argument for Woodruff's use would be appropriate. None is offered.

Woodruff's account also gives rise to a problem about the relationship between groups and their individual members. In order to accept his account, we must construe a pattern of insults to individuals on account of their membership of a group as also being insulting to the group. If we make such a move, then plainly it will be because we think it justified by some feature of insults, rather than because we view the relationship between a group and its members as being such that whatever befalls an individual on account of group membership, e.g., a punch on the nose, also befalls the group. Let us call the relevant feature, which is present in insults but absent from punches on the nose, 'distributability'. The difficulty in Woodruff's account arises when he speaks of compensatory discrimination, for he holds that 'every member is benefited equally by an act that tends to break the pattern' (p. 160). But of what benefit is it to someone who has no interest in banking that, as a result of compensatory discrimination, banking positions are now more accessible to members of his group? Surely none, unless we make the fanciful assumption that acts of compensatory discrimination are some kind of group compliment and hence "distributable". The most that can be said is that if such compensatory acts really do destroy the pattern, then those members of the group without interest in banking are no longer insulted. But ceasing to be insulted differs from being compensated for an insult.

Woodruff's account goes astray, I suggest, because be believes a morally neutral definition of discrimination to be possible, and is thus led to ask what is wrong with being discriminatory. He equates discrimination with according people 'differential treatment on morally irrelevant grounds'. He argues that this is not always wrong. Indeed. If I understand correctly what he means by 'morally irrelevant', such treatment frequently occurs without anyone suggesting that discrimination is being practised. Whether a customer patronises a particular retailer may depend on his wish to make friends with one of the retailer's employees. A woman in the happy position of being able to choose between two prospective spouses may prefer a senile millionaire to someone who would make her a better husband. In war a soldier with time for only one shot at an enemy group may opt for the only man with a beard as his target. All is fair in love and war.

What these cases have in common is that none of them can plausibly be considered as breaches of an obligation to be fair. They are either not

unfair or drawn from areas where fairness does not matter. In general it is only appropriate to describe a practice as discriminatory if we think that an obligation of fairness is breached. That Woodruff can claim that there is nothing wrong with isolated acts of discrimination in employment is, I suggest, due to the fact that he adopts the terminology of those who think it is unfair: as Hare might put it, his use of the term 'discrimination' is in inverted commas.

The light this casts on compensatory discrimination depends on how we rate obligations of fairness. For those who think that such obligations are absolute, compensatory discrimination is always wrong. Others who agree that obligations of fairness take precedence over other types of obligation will allow the possibility that we may sometimes have to choose between being fair to different groups. For them, compensatory discrimination cannot be excluded but must be assessed on the merits of the particular case. The same will be true for those who believe that obligations to promote utility sometimes take precedence over obligations of fairness.

Those who would be prepared to give weight to utilitarian considerations may be prepared to support discrimination in favour of a group, e.g., children raised in slum conditions, without needing to be convinced that the group had been wronged. This may also be true of those concerned with competing considerations of fairness. Extending special opportunities to poor children may, even if child poverty is simple misfortune, and even if some would surmount the handicap without special help, be thought fairer than the alternatives. For these reasons, though I have for convenience followed Woodruff's terminology, it would be better to speak of positive, rather than compensatory, discrimination.

University of Stirling   © STANLEY S. KLEINBERG 1976

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; Margaret A. Boden, A. E. Fisher Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. **Articles submitted for publication should be addressed to Christopher Kirwan, Exeter College, Oxford OX1 3DP.** Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

1 5 NOV 1977

ANALYSIS

Edited by
CHRISTOPHER KIRWAN

CONTENTS

ANALYSIS COMPETITION "PROBLEM" NO. 16

THE sixteenth problem is set by Professor R. J. Butler of the University of Kent at Canterbury, with acknowledgements to Mr. Robin Taylor, and it is as follows:

> If Brown in an ordinary game of dice hopes to throw a six and does so, we do not say that he threw the six intentionally. On the other hand if Brown puts one live cartridge into a six-chambered revolver, spins the chamber as he aims it at Smith and pulls the trigger hoping to kill Smith, we would say if he succeeded that he had killed Smith intentionally. How can this be so, since in both cases the probability of the desired result is the same?

Entries (of not more than 600 words) should be sent to the Editor of ANALYSIS by the 31st August, 1977. They should be accompanied by a stamped addressed envelope or international postage coupon, if return of the typescript is desired. No entries should be sent to Professor Butler. Contributors may submit entries under their own names or a pseudonym. Contributors must be under the age of thirty, or undergraduates or graduate students.

A report with any winning entries will be published in volume 38 of ANALYSIS. The ANALYSIS Committee has voted a sum of £40 which will be awarded as a prize if the adjudicator finds a sufficiently deserving contribution.

It is hoped to publish the report on Problem No. 15 in the next issue.

PLANTINGA AND THE ACTUAL WORLD

By Michael J. White

IN Sections 4 and 5, Chapter IV, of his book *The Nature of Necessity*, Alvin Plantinga condemns a certain 'confusion' he believes to be generated from David Lewis's analysis of the phrase 'the actual world' and related uses of 'actual'. Lewis analyses 'actual' and its cognates as indexical expressions. For example, the noun phrase 'the actual world' "refers at a world w to the world w" (David Lewis, 'Anselm and Actuality', *Nous*, vol. IV no. 2, May 1970, p. 185). 'Actual', Lewis maintains, is closely analogous to the indexical 'present', which "refers at any time t to the time t" (*ib.*).

Plantinga, in the sections mentioned above, takes issue with those who, following Lewis's analysis, have concluded that the term 'actual' does not "significantly or importantly" distinguish this world α, in which we live and move and have our being, from some (any) other possible world β: for it is just as true *in* β, according to the indexical analysis of 'actual', that β is the actual world as it is true *in* α that α is the actual world. Plantinga, however, claims that the reasoning of those who claim that "α's being actual does not significantly distinguish it from these other worlds . . . is confused" (Alvin Plantinga, *The Nature of Necessity*, Oxford 1974, p. 48).

I wish to argue that Plantinga's own conception of the actual world is equivocal. His analysis suggests that he conceives of the phrase 'the actual world' both as an indexical expression (and thus as an expression the denotation of which depends on its context of use) and, inconsistently, as an expression that always denotes the same world, whatever its context of use. If, however, Plantinga is interpreted as analysing a semantically simple predicate, 'is the actual world', rather than a denoting term, an analogous ambiguity in the concept of the property expressed by this predicate appears. I analyse Plantinga's claims and arguments, and various plausible interpretations of them, in considerable detail because this exercise, it seems to me, forcefully points up a deeply seated ambiguity in the conception of the actual world shared by many philosophers.

Some people, Plantinga contends, have mistakenly concluded that the phrase 'the actual world' and the phrase 'this world' are synonymous or, "less sweepingly", have concluded that the following two sentences, (1) and (2), express the same proposition.

(1) This [world] is the actual world.
(2) This world is this world.

It is against the latter "less sweeping" conclusion that Plantinga produces an argument. Suppose, for the sake of a *reductio*, that sentences (1) and (2) do, presumably when employed in this world α in which we live and move and have our being, express the same proposition. The proposition expressed by (1), Plantinga points out, is at least logically equivalent to that expressed by the following sentence (3), where 'α' is presumably non-indexical:

(3) α is the actual world.

But, according to Plantinga, the proposition expressed by sentence (3), and hence that expressed by (1), is contingent, while that expressed by (2) is not. We have our contradiction and can conclude that our original premiss, that sentences (1) and (2) express the same proposition, is false.

, So far, perhaps, so good. However, Plantinga proceeds to make two additional claims:

(A) Sentence (1) "could not have been used to express a false proposition; that is, for every world W, the proposition [(1)] expresses in W is true in W" (Plantinga, *op. cit.* p. 50).

(B) The proposition sentence (1) *does* express in α—"the one it expresses in *this* world"—is not true in every world W (*ib.*).

If sentence (1) is to express a true proposition in each world in which it is employed and if we treat—as Plantinga apparently does—the phrase 'this [world]' as an indexical expression the denotation of which depends upon the world in which it is employed, the following analysis of the sentence is plausible. The sentence expresses an identity relation and, consequently, the phrase 'the actual world' must also be treated as an indexical, the denotation of which varies with the 'context' (the world in which it is employed) *in exactly the same way* as does the denotation of 'this world'. Thus, in world α, the proposition expressed by sentence (1) will be the same as that expressed by 'α is α' (where 'α' is non-indexical); in world β sentence (1) will express the same proposition as that expressed by 'β is β', etc. Plantinga's claim (A) will be correct under this sort of indexical analysis of the phrase 'the actual world' as it occurs in sentence (1). The preceding analysis is perhaps the most straightforward analysis of sentence (1) which verifies claim (A).

However, if this indexical analysis of (1) is correct, Plantinga's claim (B) seems wrong. The proposition sentence (1) does express in this world α, in which we live and move and have our being, is the same as that expressed by 'α is α', where 'α' is non-indexical. This proposition, one of straightforward identity, surely is true in all possible worlds. Even if Plantinga were to adopt an analysis of 'identity sentences' according to which such a sentence does not express a true proposition in worlds in which the terms flanking the identity predicate do not denote (e.g., 'Pegagus is Pegasus' in this world α), our 'α is α' is not an example of such a sentence. The name 'α', according to Plantinga, *does* denote in all worlds, when employed in worlds other than α as well as when employed in α. "Each world exists in each world", as Plantinga puts it (in an admittedly Plotinian fashion) (Plantinga, *op. cit.* p. 47). Thus, if we adopt the indexical analysis of sentence (1), an analysis that makes claim (A) true, claim (B) must be false: the proposition expressed by sentence (1) in this world (identical to that expressed by the sentence 'α is α', where 'α' is non-indexical) *is* true in every possible world.

More important, however, is the relation between the indexical analysis of the phrase 'the actual world' that verifies claim (A), and Plantinga's argument that sentences (1) and (2) do not express the same

proposition. Recall that Plantinga had argued that the proposition expressed by (3) at α, and thus the equivalent proposition expressed by (1) at α, are presumably contingent. Hence they cannot express the same proposition as that expressed by (2), presumably also at α, which he says is not contingent. However, our indexical analysis indicated that in order for Plantinga's claim (A) to be correct, the proposition expressed by sentence (1) at α is the same identity proposition as that expressed by 'α is α', where 'α' is non-indexical. Furthermore, we argued that, owing to his view of possible worlds, Plantinga must hold that identity sentences about possible worlds express genuine identity relations, relations which, if they are true, are true in all possible worlds. Thus the proposition expressed by sentence (1) at α is necessary, not contingent. It follows that Plantinga cannot argue, as he must in order to obtain his *reductio*, that sentences (1) and (2) express different propositions *because* the proposition expressed by the former sentence is necessary, that expressed by the latter sentence contingent. If Plantinga is to save his claim (A), he apparently must sacrifice his argument that sentences (1) and (2) do not express the same proposition.

One possible source of confusion in Plantinga's argument pertains to sentences (1) and (3). As Plantinga himself implies, they express the same proposition in this world α, in which we live and move and have our being. The two sentences, however, are not synonymous in the sense of expressing the same proposition in each possible world. According to the indexical analysis of the phrase 'the actual world' suggested by Plantinga's claim (A), sentence (1) will express, in world β, the same proposition as that expressed by 'β is β', where 'β' is non-indexical, while sentence (3) will express the same proposition as that expressed by 'α is β'. Briefly, in each world sentence (1) will express a necessarily true (identity) proposition; sentence (3), however, will express a necessarily true (identity) proposition only in α, and a false proposition in worlds other than α. It is perhaps this fact that suggests to Plantinga the idea— inconsistent with his claim (A), I have argued—that the proposition expressed by sentence (3) *in fact* (i.e., at α, where we live and move and have our being) is contingent.

I should want to argue that, given the indexical interpretation of the phrase 'the actual world' suggested by claim (A), sentences (1) and (2) are synonymous in the sense of expressing identical propositions at each possible world (not, of course, the same proposition from world to world). Furthermore, I should want to claim that, despite what Plantinga says, the phrases 'the actual world' and 'this world' are synonymous in the same sense: in each possible world they have the same denotation, namely, that possible world. Since the argument by means of which Plantinga hopes explicitly to falsify the former claim (and implicitly to falsify the latter claim, I believe) is inconsistent with the indexical

analysis of his claim (A), the argument does not, under this interpretation, accomplish its purpose. Plantinga seems to waver between a conception of the phrase (as in claim (A)) that suggests Lewis's indexical analysis and a conception of the phrase as a non-indexical expression singling out a unique entity, whatever the context of use of the expression.

There is, however, another analysis that interprets sentence (1) *not* as expressing an identity relation between the referents of two denoting terms but as predicating a property of a world. The syntactically complex predicate 'is the actual world' is treated as a single semantic unit; this predicate's 'semantic value' is a property Φ, roughly the property of being-the-nitty-gritty-world-in-which-we-*really*-do-live-and-move-and-have-our-being (this interpretation was suggested to me by David Blumenfeld). Then, sentence (1), in this world α, expresses the same proposition as that expressed by '$\Phi(\alpha)$' or 'α has property Φ'. In world β, sentence (1) expresses the same proposition as that expressed by '$\Phi(\beta)$' or 'β has property Φ', etc.

Does this analysis allow Plantinga to save both claims (A) and (B) and his *reductio* argument? The first question that arises concerning this new predicate-interpretation of sentence (1) is whether it verifies claim (A). Let us first consider the property Φ, the property of being-the-actual-world, *simpliciter*. It is fairly obvious that Plantinga wants this property *really* to belong to some worlds (more precisely, one world, α), and not all. Hence, the proposition expressed by 'α has property Φ' is true, but those propositions expressed by 'β has the property Φ', 'γ has property Φ', etc. are false. However, this analysis indicates that claim (A) is not true: sentence (1) does express false propositions at some worlds. For example, the proposition that it expresses at β would be that expressed by 'β has property Φ', and we have just said that this proposition is false.

We can save claim (A) only by fiddling either with our property Φ or with our conception of the nature of the truth of propositions about possible worlds. We might substitute for Φ a set of 'world-indexed' properties one (and only one) of which each world trivially possesses (see the discussion of world-indexed properties in Plantinga, *op. cit.*, pp. 62ff). Thus, α has the property of being-actual-at-or-in-α, β has the property of being-actual-at-or-in-β, etc. Supposedly, sentence (1), when employed at α, will pick out the proposition expressed by 'α has the property of being-actual-at-α', when employed at β, will pick out the proposition expressed by 'β has the property of being-actual-at-β', etc. Thus, given our understanding of these rather peculiar properties, sentence (1) will not fail to express a true proposition, whatever world it is employed in. However, Plantinga maintains that all propositions attributing such world-indexed properties to something are either

necessarily true or necessarily false. Since, *ex hypothesi*, sentence (1) expresses a true proposition at α, it does *not* here express a contingent proposition, and Plantinga's *reductio*, which depends on the claim that (1) 'in reality' (i.e., at α) expresses a contingent proposition, fails.

What initially seems a better way to proceed begins with emphasizing Plantinga's distinction between truth *simpliciter* and truth-at-W, where W is any possible world (Plantinga, *op. cit.* p. 55). Φ, though a property properly attributed only to possible worlds, is not world-indexed in any sense. However, depending upon the world from which one is 'viewing' or considering the set of possible worlds, the property belongs to different worlds. The extension of the property is thus dependent upon this world 'viewpoint'. If I am viewing the set of possible worlds from α, the property belongs only to α; if I am viewing the set of worlds from β, the property belongs exclusively to β, etc. Hence it follows that claim (A) is correct: for every possible world W, the proposition sentence (1) expresses in W (i.e., the proposition that W has property Φ) is true-in-W. Claim (B) is also preserved: the proposition that (1) expresses in α, the proposition that α has property Φ, is not true in other worlds because from the viewpoint of world W ($W \neq α$), W has property Φ but α does not. Hence, the proposition expressed by sentence (1) at α is contingent, and Plantinga's *reductio* argument, which depends on this fact, succeeds.

A lingering doubt may remain as to whether this analysis does not sacrifice the 'special status' Plantinga wishes to accord to this world α, in which we live and move and have our being. Plantinga has a handy answer: there is a big difference between truth-in-W, for any possible world W, and truth *simpliciter*. It may be true-in-β that β is the actual world. However, it no more follows from this fact that it is true *simpliciter* ('really true') that β is the actual world than it follows from the fact that it is true-in-β that Socrates fled his prison in Athens that it is true *simpliciter* that Socrates fled his Athenian prison. The 'special status' of α, where we live and move and have our being, is preserved. It may be true-in-β that β has the property of being-the-actual-world, but it is nonetheless not true *simpliciter*, i.e., it is false, that β has this property. It is only of α that it is true *simpliciter* that it has this property, the property of being-the-actual-world. Plantinga apparently can accomplish a great deal with the aid of the true-*simpliciter*/true-in-W distinction: he can preserve claim (A); he can preserve claim (B); he can save his *reductio* argument that sentences (1) and (2) do not express the same proposition; he can at least implicitly establish that the phrases 'this world' and 'the actual world' are not synonymous; finally, he need nowhere equivocate on the predicate 'is the actual world', everywhere assigning it the *same* property Φ, as just analysed.

I wish to argue, however, that this picture of success is deceptive. The analysis that produces it depends on the crucial distinction between

truth *simpliciter* and truth-in-W, for any possible world W. Unfortunately, the distinction seems to rest on something like the distinction between the actual world and other worlds that are merely possible but not actual. A proposition is true *simpliciter* if and only if it is true in the actual world, while it is true-in-W if and only if it is true in possible world W, whether W is actual or not. To suppose that truth *simpliciter* is relative to the world from which one imagines himself to be considering a question of truth, i.e., relative to one's possible world 'viewpoint', is to collapse the distinction between truth and truth-in-W and, ultimately, to collapse the distinction between truth ar d possibility. Even if I were 'to assume the viewpoint' of possible world β, in which it is true that Socrates fled his imprisonment, it nonetheless is only true-in-β, not *actually* true or true *simpliciter*, that Socrates fled. The point is that the concept of the *real* or *actual* world that underlies the true-*simpliciter*/true-in-W distinction is that of some *one* world's possessing the distinction of being the actual or real world. The property Φ (the property of being-the-actual-world and thus the world by means of reference to which truth *simpliciter* is determined) is not relativized to one's 'world viewpoint'. In order to get the desired concept of truth *simpliciter* we must pick out the one actual or real world 'from above', as it were. To advance the idea that we are a 'prisoner' of the possible world we just 'happen' to occupy and that other possible worlds, being 'actual' from their own viewpoint, are thus equally referents for truth *simpliciter* is to forget that the notion of a possible world is supposed to be a semantic tool, not a semantic stumbling block.

However, as we saw, this last analysis of the property Φ, the property of being-the-actual-world, makes the extension of this property world-dependent. From the viewpoint of α, α has the propeity of being the actual world; from the viewpoint of β, β has that distinction. The result is that this last analysis, which held out the hope of reconciling Plantinga's claims (A) and (B) and his *reductio* argument, equivocates between two senses of the predicate 'is the actual world'. The *explicit* analysis of the predicate assigns it a property the extension of which varies with one's world viewpoint. The *implicit* conception of the actual world underlying the true-*simpliciter*/true-in-W distinction—by means of which we had hoped to safeguard the 'special status' of this world α in which we live and move and have our being—is a different conception, one that assumes a 'meta-world viewpoint' and satisfies the following principle:

$$(\exists \chi)((\chi \varepsilon \{\psi : \psi \text{ is a possible world}\}) \text{ and } (\chi \text{ is the actual world}) \text{ and }$$
$$(\phi) (\phi \text{ is the actual world} \supset \phi = \chi)).$$

The outcome of the attempt to save Plantinga's claims (A) and (B) and his *reductio* argument by interpreting sentence (1) as predicating a property of a world rather than as expressing an identity relation

between worlds results in an ambiguity analogous to the indexical/non-indexical ambiguity of the phrase 'the actual world', when sentence (1) is interpreted as expressing an identity relation. Neither interpretation renders Plantinga's position consistent.

Although Plantinga's claims (A) and (B) and his *reductio* only appear consistent because of equivocation, the intuitions underlying his position are appealing. This fact may well suggest that our concept of the actual world is deeply ambiguous. I find myself in much the same ambivalent frame of mind that Plantinga manifests. Neither the indexical analysis of the phrase 'the actual world' (or the analogous, context dependent interpretation of the predicate 'is the actual world') nor the non-indexical analysis (or the analogous, 'meta-world' interpretation of the predicate) seem separately both to satisfy all our linguistic intuitions and to meet the requirements imposed by the use of this concept in 'doing' semantics.

Arizona State University   © MICHAEL J. WHITE 1977

PRIOR'S THEORY OF PROPOSITIONS

By PHILIP HUGLY *and* CHARLES SAYWARD

A N. Prior has propounded a theory which, if correct, explains how it is possible for statements about propositions to be true even if propositions do not exist. The major feature of his theory is its treatment of sentence letters as bindable variables in non-referential positions. His theory, however, does not include a semantical account of the resulting quantification. We here take some steps toward filling this gap.

I. PRIOR'S THEORY

Consider the following two examples of sentences about propositions:

(1) John believes that man is mortal.
(2) John believes something.

On the surface (1) is of the form 'Rab'. Believing is construed as a relation which holds between John and a proposition if (1) is true. Homogeneity of treatment yields

> There exists an object x such that x is a proposition and John believes x

as an analysis of (2).

But there is reason for thinking this surface analysis to be wrong, for serious utterances of sentences like (1) and (2) are not normally understood to commit their utterers to the existence of propositions. But what, then, should replace the surface analysis?

Attempts to reduce talk about propositions to talk about something else, e.g. sentences, seem doomed from the start. For a statement like

(3) There are truths about real numbers which will never be expressed in any language

seems bound to resist any such attempt.

Prior's theory purports to explain how statements like (1)–(3) can be true if propositions do not exist and talk about propositions does not reduce to talk about language.

Basic to the surface analysis is the idea that 'believes', 'asserts', 'doubts', and the like are binary operators forming sentences from a pair of names or singular terms. Prior denies this and instead treats 'John believes that' as a unary operator similar to 'it is not the case that' in forming a sentence from a sentence. This operator can in turn be viewed as a binary operator which forms a sentence from a name and a sentence (it is like a predicate at one end and a connective at the other). Thus (1) can be viewed as of the form 'x believes that p', where 'x' stands in for a name and 'p' stands in for a sentence. This is in sharp contrast to the surface analysis which would lead to the form 'x believes y' with both 'x' and 'y' standing in for names or namelike expressions.

The surface analysis thus represents the position of 'y' as referential. But not so for the position of 'p' on Prior's analysis: for 'p' stands in for a sentence and a sentence is not a term of reference. Coupled with this is the idea that sentence letters are as much subject to quantification as are name letters. It is this idea that Prior uses in handling statements that ostensibly quantify over propositions. Shortly put, the idea is to quantify with respect to sentence letters, which thus become genuine variables and yet stand in non-referential position. Thus

(4) For some p, John asserted that p

is Prior's analysis of (2).

II. THE THEORY'S GAP

It seems that the only way to account for the truth conditions of a statement like (4) is via the substitutional account of quantification. For, since 'p' is a sentence variable and not in referential position, no sense can be made of (4) using the referential account, and it and the substitutional account are the only accounts going. But now consider how (3) would be dealt with on Prior's analysis:

> For some p, it is true that p and that p is about real numbers and
> that p will never be expressed in any language.

It clearly will not do to say this sentence is true in English (or, perhaps,
a symbolic variant of English) just in case some substitution instance of
its unquantified part is true in the language. For such reasons Prior
disavows the substitutional account: there is no more need that a
language express every proposition than that it name every red haired
object (A. N. Prior, *Objects of Thought*, edited by P. T. Geach and
A. J. P. Kenny, Oxford University Press, 1971, p. 36).

The upshot is that there is a serious gap in Prior's theory. We know
how '$\Pi p \ldots p \ldots$' and '$\Sigma p \ldots p \ldots$' are *not* to be interpreted; but this
does not tell us how they *are* to be interpreted. This can be put less
abstractly by considering the formal language Prior describes in *Objects
of Thought*, p. 101:

> ... (a) ordinary propositional calculus enriched with variables for
> expressions which form sentences from sentences, with quantifiers
> binding variables standing for sentences, and with an identity
> function with sentences as arguments; (b) the ordinary theory of
> quantification applied to our special quantifiers; and (c) ordinary
> laws of identity applied to our special function.

Prior makes clear the general outlines of a syntax of this language, but
gives only hints as to its semantics. The situation is analogous to what
one encountered when presented with a language for quantified modal
logic specified by formation and transformation rules but lacking a
semantics. In this case the gap was filled by defining (e.g., as was done
by Kripke) a model for the language. In the case of a Priorean language
the gap is to be filled in the same way.

III. THE SYSTEM E

Two desiderata motivate the construction of a semantics that is in
accord with Prior's theory. First, there must be at least one model M
such that a quantification '$\Pi p \ldots p \ldots$' is false in M even though all of its
instances are true in M; i.e. the semantics must be non-substitutional.
Second, the semantics must provide no basis for saying the positions of
the bound sentential variables in a quantified sentence are referential;
i.e. the semantics must be non-referential.

We now construct a very simple language E and argue that a semantics
can be given for E that satisfies both of the conditions just mentioned.

The Syntax of E

1. The vocabulary of E:

   $\Pi, p, P,$ '

2. The sentential variables:

$p, p', p'' \ldots$

3. The sentential constants:

$P, P', P'' \ldots$

4. The sentences of E:

    (i)   all sentential constants;
    (ii) Παα, if α is a sentential variable;
    (iii) nothing else.

5. Where Γ is a set of sentences of E and φ is a sentence of E, a deriva-
tion in E is a set of ordered pairs <Γ, φ> constructible in accordance
with the following rules:

    P: → <{φ}, φ>.
    US: <Γ, Παα> → <Γ, β>, where β is a sentential constant.
    UG: <Γ, β> → <Γ, Παα>, if the constant β is not in any
    sentence of Γ.

6. φ is derivable from Γ if and only if there is a derivation of which
<Γ, φ> is a member.

7. φ is a thesis of E if and only if φ is derivable from the null set.

The Semantics of E

8. A domain D is a set such that x ε D just in case

    (i)   x is a sequence of natural numbers, or
    (ii) x={y: y ε D and y is a sequence of natural numbers}.

9. A sequence function S is a function from the sentences of E to D,
which satisfies these conditions:

    (i)   if β is a sentential constant, S(β) is a sequence of natural
    numbers;
    (ii) if α is a sentential variable, S(Παα)={S'('P'): where S'
    differs from S at most in what it assigns to 'P'}.

10. A valuation function V is a function from D to {1, 0}, which satisfies
these conditions:

    (i)   if x is a sequence of natural numbers and a member of D,
    V(x) is 1 or 0 but not both;
    (ii) V(S(Παα))=1, if, for all x ε S(Παα), V(x)=1; V(S(Παα))=0
    otherwise.

11. A model for E is an ordered triple <D, S, V>, where D is a domain,
S is a sequence function and V is a valuation function.

12. A sentence ϕ of E is true in a model $<D, S, V>$ just in case $V(S(\phi)) = 1$; ϕ is a valid sentence of E just in case ϕ is true in every model for E; ϕ is a consequence of a set of sentences Γ just in case, for any model, ϕ is true in that model if each sentence of Γ is.

It is easy to show that a sentence of E is a thesis just in case it is valid. The semantics of E is not substitutional. The biconditional

> $\Pi\alpha\alpha$ is true in $<D, S, V>$ if and only if every sentential constant is true in $<D, S, V>$

is true for some models only. Call the elements designated by 8 (i) 'the basic elements of D'. The biconditional is true for those models in which every basic element of D is assigned by S to a sentential constant. But when not every basic element of D is so assigned, $\Pi\alpha\alpha$ might be false while each sentential constant is true. Here is an example. Let the basic elements of D be the set of all sequences of natural numbers. (In this case D is indenumerable, so that no function assigns each element of D to a sentential constant.) If β is the ith sentential constant of E, $S(\beta) = <i>$. If x is a basic element of D, $V(x) = 1$ if, for some constant β, $x = S(\beta)$; otherwise $V(x) = 0$. Each sentential constant is true in this model, but 'Πpp' is false.

IV. System E and Referential Quantification

The semantics of E is not referential. We defend this claim by replying to the most plausible objection to it that we can think of.

The objection goes as follows: D is a set of *objects*. Function S assigns an object to each sentence of E. Thus each sentence is semantically construed as a name of its assigned object, as *denoting* that object. But then sentences are, after all, construed as standing in referential position, and so also for the variables which replace them. The semantics provided for E, so far from providing an account of the binding of variables in non-referential position, simply provides a standard referential account of quantification. The semantic reality underlying the syntactic appearance is that of reference.

This objection is faulty. From the fact that S assigns an object to a sentence ϕ it does not follow that ϕ names that object. From, e.g., '$S('P') = <3>$' one cannot conclude ' 'P' names $<3>$'.

A comparison between the semantics of E and the semantics of an ordinary first order language is relevant. For a particular function V for the latter we may correctly write '$V('a') = $ John' and '$V('Fa') = 1$'. Similarly, where 'R' designates the positive square root function we may correctly write '$R(4) = 2$'. Now, in this case it is correct to say that '$R(4)$' names the number 2. But it would clearly be incorrect to say that 4 names the number 2, for 4 is not even a name. We can also say that '$V('a')$'

names John. But from this alone, as the above example shows, it does not follow that 'a' names John. To obtain this result we would need the further stipulation that 'a' is to name (denote) V('a'). Such stipulation is commonly taken for granted in connection with individual constants like 'a'. But nothing like this is or need be involved in the case of 'Fa'. The situation here is like that for the positive square root function on 4. 'V('Fa')' indeed names the number 1. But it would clearly be incorrect to say that 'Fa' names that number, for 'Fa' is not even a name. '$Fa=1$' is as illformed as is 'John is happy equals one'. In sum, it would be a mistake to think that a sentence of a first order language is treated as a name of 1 or 0 just because an evaluation function assigns 1 or 0 to it. It would be equally mistaken to think a sentential constant names an element of some domain just because S assigns it that element. Thus the semantics of E does not require construing the sentential variables of E to be name variables.

Prior has insisted that one does not have to choose between the referential and substitutional accounts of quantification to make sense of, e.g., 'for all p it is true that p'. The semantics of E bears him out on this.

V. PROPOSITIONAL IDENTITY

E is a language of minimum deductive power. We have made it so for the sake of avoiding technicalities that would tend to obscure our basic result, which is that a semantical account of quantification that is in accord with Prior's theory *can* be given. If this idea is accepted it is worth exploring extensions of E. Of particular interest is what results from adding to E the identity function with sentences as arguments. For such a language, unlike E, will be nonextensional.

Changes in the basic syntax. To the vocabulary of E we add the symbol 'I'. The constants and variables will be as before. A sentence of the enlarged system (call it 'F') will be a formula of F that does not contain a free occurrence of a variable. (We assume the usual understanding of free and bound occurrences of variables.) The formulas of F then are:

    (i)   all sentential constants and variables;

    (ii) $\Pi\alpha\psi$ if α is a variable and ψ a formula, and α has at least one occurrence in ψ which is in no part of ψ of the form $\Pi\alpha\chi$;

    (iii) $I\phi\psi$ if ϕ and ψ are formulas;

    (iv) nothing else.

For technical reasons we wish to exclude vacuous quantifications from the set of formulas. This occasions the 'if' part of (ii).

Changes in the rules of inference. The changes in the quantifications brought about by (ii) above require these two changes in US and UG.

US: $<\Gamma, \Pi\alpha\psi> \to <\Gamma, \psi\alpha/\beta>$
UG: $<\Gamma, \psi\alpha/\beta> \to <\Gamma, \Pi\alpha\psi>$ if the constant β is not in any
   sentence of Γ or in the formula ψ.

Following customary usage '$\psi\alpha/\beta$' means 'the result of replacing each
free occurrence of the variable α in ψ by the constant β'.

According to Prior the usual laws of identity are to govern 'I'.
Following this idea

Id(i): $\to <$the null set, $I\phi\phi>$ where ϕ is any sentence,
Id(ii): $<\Gamma, I\phi_1\phi_2> \to <\Gamma, I\psi_1\psi_2>$ where ψ_2 is a sentence that
results from the sentence ψ_1 by replacing some occurrence of
sentence ϕ_1 in ψ_1 by the sentence ϕ_2,

would be the only rules of inference governing propositional identity in
F. But it is clear to us that these two rules are not sufficient.

For one thing the converse of Id(ii) should hold for propositional
identity. For example if $IIPQIRQ$ then IPR. (For Prior this is a per-
spicious way of saying that if the proposition that P and Q are identical
propositions is itself identical with the proposition that R and Q are
identical propositions then the proposition that P is identical with the
proposition that R). So a third identity rule, Id(iii), that is needed is
just the converse of Id(ii).

Secondly, a plausible thing to hold of propositional identity is that,
e.g., '$\Pi p'p''$ and 'Πpp' say the same thing.

Id(iv): $\to <$the null set, $I\phi\phi\alpha//\alpha'>$

'$\phi\alpha//\alpha'$' means 'the result of replacing each occurrence of a variable α in a
sentence ϕ by an occurrence of a variable α' that does not occur in ϕ'.

Next a plausible truth about propositional identity is $IIPQIQP$.
This can be generalized. Let $\bar\phi$ be the result of replacing some occurrence
of a component sentence $I\psi_1\psi_2$ of ϕ by $I\psi_2\psi_1$. Then $\bar\phi$ and ϕ say the same
thing.

Id(v): $\to <$the null set, $I\phi\bar\phi>$

Finally, note that from $I\phi_1\phi_2$ and ϕ_1 we would want to infer ϕ_2.

Id(vi): $<\Gamma, \phi_1>, <\Delta, I\phi_1\phi_2> \to <\cup\Delta, \phi_2>$

(This rule would be dispensable if we added to F the truth-functional
connectives.)

The semantics of F. Under what conditions will $I\phi\psi$ be true in a model?
Just in case ϕ and ψ have the same propositional content. In our termin-
ology, just in case $S(\phi)=S(\psi)$. So the rule we want is

$V(S(I\phi\psi))=1$ iff $S(\phi)=S(\psi)$.

In the case where β is a sentence constant $V(S(\beta))$ is defined as before, as is $V(S(\Pi\alpha\psi))$ for variable α and formula ψ.

What must ϕ, ψ, S and D satisfy in order that $S(\phi)=S(\psi)$ in a model $<D, S, V>$?

This question is best approached by first defining the following relation in the class of formulas of F:

> (i) Any two variables are quasi-equiform and any two constants are quasi-equiform.
>
> (ii) If $\phi=I\chi_1\chi_2$ then ϕ is quasi-equiform with ψ just in case, first, $\psi=I\chi_3\chi_4$, and also secondly χ_1 is quasi-equiform with χ_3 and χ_2 is quasi-equiform with χ_4 *or* χ_1 is quasi-equiform with χ_4 and χ_2 is quasi-equiform with χ_3.
>
> (iii) If $\phi=\Pi\alpha\chi$ then ϕ is quasi-equiform with ψ just in case $\psi=\Pi\alpha'\chi'$ and χ is quasi-equiform with χ'.

Among the many decisions regarding propositional identity the following seems at least as reasonable as any other, particularly in view of the syntax of F: For any model $<D, S, V>$ and any sentences ϕ, ψ, $S(\phi)=S(\psi)$ only if ϕ and ψ are quasi-equiform sentences.

The above decision imposes certain constraints on a construal of the set of domains D and sequence functions S suitable for F. First, D will have to consist of three different sorts of elements: elements that get assigned to the constants, elements that get assigned to the identities and elements that get assigned to the quantifications. Second, we want

> $\Pi\alpha\phi$ is true in $<D, S, V>$ iff $\phi\alpha/\beta$ is true in $<D, S, V>$ for every constant β

to be false for some models to insure that the quantification in F is not substitutional. Finally, for any D at all, it must be the case that, for any sentence ϕ, $S(\phi)\ \varepsilon\ D$.

Here's the definition. A domain D is a union of sets D_0, D_1, . . ., D_n, . . . satisfying the following two conditions:

> (i) D_0 is a set of at least 3 sequences of natural numbers.
>
> (ii) For each $i>0$, D_i is a set of objects x satisfying one of these three conditions:
>
> > (a) $x\ \varepsilon\ D_{i-1}$;
> >
> > (b) for some y, z $\varepsilon\ D_i$, $x=\{y, z\}$;
> >
> > (c) for some $X\subseteq D_{i-1}$, $x=X$.

A sequence function S will be understood (as before) to map sentences of the language onto elements of D. In the case of F such an S will satisfy these conditions:

(

(i)   for any constant β, $S(\beta) \in D_0$;

(ii) for any sentences ϕ and ψ, $S(I\phi\psi) = \{S(\phi), S(\psi)\}$;

(iii) for any formula ψ with at most free occurrences of variable α, $S(\Pi\alpha\psi) = \{S'(\psi\alpha/\beta): \beta$ is the first constant not in ψ and S' differs at most from S in what is assigned to $\beta\}$.

Thus, constants get assigned members of D_0; identities get assigned sets with one or two members; quantifications get assigned sets with three or more members.

Final remarks. F is a nonextensional language because '*I*' is a sentential connective that is not truth-functional: The truth value of $I\phi\psi$ in a model is not a function of the truth values of ϕ and ψ in that model. E.g., there are models for which both

$$V(S(\phi)) = V(S(\psi)) = V(S(\phi')) = V(S(\psi'))$$

and

$$V(S(I\phi\psi)) \neq V(I\phi'\psi'))$$

hold. Also quantification in F is interpreted neither as substitutional nor as referential. The justification of both claims is virtually the same as was given in the case of E.

Thus F is a language in accord with Prior's syntactic insights and supplied with a semantics such that (i) F contains nonextensional contexts and (ii) the quantifications in F are neither substitutional nor referential.

Granting this, further questions may now be raised. E.g., are the rules of inference for F sound? Are they complete in the sense that a sentence of F is valid only if it is a thesis? If the inference rules for F are not in that sense complete, can further provably sound rules be added so as to insure that kind of completeness?

We here answer none of these questions. For some the answers depend upon relatively simple proofs. Others depend upon proofs of a significant degree of difficulty. The aim of this paper has not been to answer such questions as these, but to provide a basis apart from which such questions could not be raised in the first place. If that basis is correct, then it goes almost without saying that there is need for further investigation.

University of Nebraska, Lincoln

© Philip Hugly *and* Charles Sayward 1977

THE CONCEPT OF THE SUPERNATURAL

By GILBERT FULMER

TRADITIONAL theism holds that the universe is dependent on a Creator who is supernatural, since he created *ex nihilo* the natural order, which is sustained only by his will. It has sometimes been thought that the existence of this being could be demonstrated by the teleological, the cosmological, and the ontological arguments. I will argue, however, that all these notions are undermined by a common logical incoherence. In the end, I suggest, it is evident that the universe cannot be dependent on a personal Creator, but is ultimately impersonal.

1. R. G. Swinburne has attempted to revive the teleological argument.[1] It is impossible to explain all natural laws scientifically, he argues, for we must eventually come to 'the most fundamental regularities', for which no explanation is known. At this point we have progressed as far as possible in explaining events naturalistically, he claims; so if we are to continue our explanatory efforts we must entertain a different kind of hypothesis: that all natural laws result from the will of a god. If confirmed this hypothesis would greatly simplify our system of explanations, for then everything would ultimately be explained by reference to the will of this god. (Swinburne does not discuss the possibility that the most fundamental facts of nature might be statistical; nor shall I, since nothing in the present argument hinges on that issue.)

But even if Swinburne's god hypothesis were correct, it could not provide the ultimate explanation which was its goal. The purpose of the hypothesis is to explain all natural laws in terms of something *other* than natural law. The explanation is not to be based on regularities of events at all, but on the free decisions of a supernatural rational agent. Obviously, then, this god must not himself depend on natural laws. That is, the god's actions must not presuppose any facts of the natural universe; for if they did, his choices would be logically less fundamental than these facts—and would not, after all, be the ultimate explanation. But explanations in terms of Swinburne's god *do* involve an appeal to a natural law. For if the god can impose his will on the world, it is a natural law that whatever he wills, occurs. That is, it is a fact of the universe that if the god wills x, then x is the case; for example, if he wills that $e = mC^2$, it is so. And this fact cannot itself be the product of the god's will; for if it were not a fact, his will could produce no effects whatever—and to make his will effective would be to produce an effect. The fact that events occur as he wills them cannot be the result

[1] 'The Argument from Design', *Philosophy* v. 43 (1968), pp. 199–212.

of his will. Thus this fact is logically more fundamental than the god's choices: his acts presuppose this fact, but not the converse.

The fact that the god's will is effective is natural in exactly the same sense that any other fact is natural: it occurs independently of any agent's choice. (Of course, it would be pointless to say that this law is due to the fiat of yet another rational agent, since it would then be a natural fact that *his* will was done.) This fact is, on Swinburne's god hypothesis, the ultimate available explanation. It is therefore impossible to explain all natural laws in animistic terms, since the very fact that a mind can effect its choices is a further law. So Swinburne's god hypothesis cannot achieve the final explanation which was to justify our entertaining it. And the reformulated design argument is as powerless as the original to establish a supernatural architect.

(A widely accepted argument for the autonomy of ethics holds that moral obligations cannot originate with divine command. For if there were no obligation to obey God's command, there would be no obligation to obey his command to obey his command; so there must be at least one obligation which does not acquire its force from divine command, namely the obligation to obey such commands. The present argument is parallel. If a being had no power to make events conform to his will, his decision could not create this power; so there must be at least one fact of nature which does not result from his will, namely the fact that his will is effective.)

2. All the elements of this argument are to be found in Sec. VII of Hume's *Enquiry Concerning Human Understanding*. And J. S. Mill stated it in so many words:

> The power of volitions over phenomena is itself a law, and one of the earliest known and acknowledged laws of nature. . . . There is, therefore, no more a supposition of violation of law in supposing that events are produced, prevented, or modified by God's actions, than in the supposition of their being produced, prevented, or modified by man's actions. Both are equally in the course of nature, both equally consistent with what we know of the government of all things by law.[1]

But its implications have not, it seems to me, received the attention they merit. Animism cannot provide an alternative to explanation by natural laws, because the power of a mind to bring about its choices is itself such a law. All that can be said, in the end, is that these are facts of nature. Failure to see this may stem in part from a lingering notion that natural laws are prescriptive, rather than descriptive. It would be unfair to charge supporters of supernaturalism with believing, *simpliciter*, that a law requires a lawgiver. But it is a similar mistake to imagine that explanations in terms of personal agency can ultimately

---

[1] 'Theism', in *Three Essays on Religion*. New York, Greenwood Press, 1969 (reprinted from the 1874 Holt edition), pp. 226–7.

account for natural laws. If a rational agent's will is always or sometimes effective, that regularity is an impersonal fact about the universe which cannot be the result of his will.

3. This has implications for the cosmological argument as well. Like the design argument, it claims that a divine artificer is required to account for observed phenomena. But, while the former argues that a designer is necessary to explain particular features of the cosmos, the latter proposes an explanation for the natural order as a whole. It is widely agreed, however, that if this argument is to succeed it must do more than merely identify a god as one more member in a causal sequence. In that case his status would be no different from that of any other cause, and it would be necessary to ask what caused the god. The only promising version of the cosmological argument, therefore, is the one which holds that the natural universe cannot be a self-supporting system of merely contingent facts, but must depend on something beyond itself: a supernatural being. It is only in this form that the argument offers a God who is the all-sustaining Creator, and the ultimate explanation of the universe.

But even if there were a bodiless rational agent who created the familiar objects and laws of nature, it would remain a natural fact that the will of that being is fulfilled. So the power of the being on which the universe is said to depend would itself be a fact of that universe. The natural order would therefore *not* have been explained in supernatural-istic terms, since there would remain a natural fact which was independent of the decision of the posited being. Thus even the strongest form of the cosmological argument cannot explain nature by reference to something which is not natural. But to do so was the only justification for introducing the idea of the creating being in the first place; so the cosmological argument can give no reason to suppose that such a being exists.

4. The present argument further shows that the doctrine of creation *ex nihilo* cannot be correct. It could not be the case that a supernatural agent created the universe including *all* the laws that govern it. For the fact that he can create would be a law of the universe, independent of his will.

5. The ontological argument can now be seen to fail for the same reason. It has been frequently argued that God's existence is either logically necessary or logically impossible. Many sceptics grant that if God existed, his existence would be logically necessary; but they insist that, because of the characteristics ascribed to him, it is logically impossible that he exist. And if the argument being urged here is sound, this must be true. For even if there were a being who created all the universe we know, it would still be a fact of nature, independent of his will, that he could put his decisions into effect. It is logically

impossible that any agent could stand above and control the whole of nature, because his very power to act would be a fact which was not the result of personal agency, and hence natural. Therefore the being himself would be a part of nature: he would be subject, as are we all, to natural law. Thus the animistic belief that nature could be the work of a supernatural Creator cannot be correct; the concept of such a being is incoherent. Whatever else the ontological argument may prove, it cannot show the existence of the God of theism.

6. So the concept of supernatural agency can provide no alternative to explanation in terms of natural facts. There must be at least one fundamental fact of the universe which is not the result of any agent's choice. And so, whatever the role of human or other personality in the universe, the universe itself must be ultimately impersonal. We must recognize this if we are to understand the world and our place in it.

Southwest Texas State University   © GILBERT FULMER 1977

WHAT VITIATES AN INFINITE REGRESS OF JUSTIFICATION?

By N. M. L. NATHAN

I

SUPPOSE that no one can simultaneously believe an infinite number of different propositions. It should be possible, on this assumption, to show that if there is any proposition which you are justified in believing only because you are justified in believing some other proposition, then there is some proposition which you are justified in believing but not because you are justified in believing some other proposition. More briefly: it should be possible to prove that if there is any proposition which you are indirectly justified in believing then there is some proposition which you are directly justified in believing. More briefly still: that indirect justification is heteronomous. I won't attempt to spell the proof out in detail, but the general idea would be this. The relation which holds between a man's being justified in believing p (i.e., the proposition that p) and his being justified in believing q, when he is only justified in believing p because he is justified in believing q, is both asymmetrical and transitive. So if we have a series of propositions such that for each member of the series, a man is justified in believing that member only because he is justified in believing some other member, then the series must be not only infinite but also non-repetitive—every

proposition in it will be different. But if no one can simultaneously believe an infinite number of different propositions, there can't be such a series; so indirect justification must be heteronomous.

Now, even if it is true that a man can't simultaneously believe an infinite number of propositions, it doesn't seem to be necessarily true. And yet it does seem to be quite widely accepted that the heteronomy of indirect justification is a necessary truth. So the question arises: can we show the heteronomy of indirect justification to be a necessary truth without also taking it as necessarily true that no one can simultaneously believe an infinite number of propositions? In other words: do we need this assumption of finitude in the objects of belief in order to show the impossibility of a regress in which one is justified in believing p only because one is justified in believing q, justified in believing q only because one is justified in believing r, and so on non-repetitively and ad infinitum?

A scrutiny of what philosophers have said against infinite justificatory regression in fact reveals only two remotely plausible arguments for heteronomy which are independent of the assumption of finitude in the objects of belief. In what follows I'll criticize them both, offer what is I hope a better and similarly independent argument for the same conclusion, and finish with a word about the bearing of this argument on scepticism about justified believing.

Everything turns, of course, on what we are to understand by 'justification', and the meaning I'll give it will for the most part have to emerge as I go on. But there is one possible misunderstanding which I should perhaps try to anticipate at the outset. I don't identify the justified believing of a proposition with the believing of a proposition for which there is sufficient evidence, or even with the believing of a proposition for which the believer has sufficient evidence, at any rate if having sufficient evidence for q is the same as believing a proposition which is sufficient evidence for p. To be justified in believing a proposition, it is necessary that you have consciously done something to assure yourself that it is at least probably true. It is necessary also that it actually is at least probably true, and if you are indirectly justified in believing it, it is necessary that you take some other proposition as evidence for it. But you must always have done something, and in the case of indirect justification doing something means investigating possible evidence. On the meaning of 'sufficient evidence' it will do for the moment to say that q is sufficient evidence for p only if (i) q is true, (ii) when propositions like q are true, so usually at least are propositions like p, (iii) the argument 'q therefore p' is not question-begging. What I mean by condition (iii) is that q and p are not so related that any normally intelligent man who was doubtful about the truth of p would be equally doubtful about the truth of q. This condition makes it incorrect to say that a proposition is

sufficient evidence for itself; cf. R. A. Jaeger, 'Implication and evidence', *Journal of Philosophy* 52 (1975), pp. 475–85. (It doesn't prevent you from saying that a proposition is self-evident, for you may then be referring to the evidentness of the proposition, rather than postulating an evidential relation which it is supposed to bear to itself; H. H. Price, *Belief* (London 1969), p. 92.)

II

The first charge against infinite justificatory regression which I want to consider is that if it were possible one could be justified in believing anything at all, including what is obviously false. Deutscher falls back on this allegation after some half-hearted doubts about the possibility of someone's believing an infinite number of propositions (Max Deutscher, 'Reasons, regresses and grounds', *Australasian Journal of Philosophy* 51 (1973), p. 6). Pollock has a more detailed development of the same theme in his recent *Knowledge and Justification*. We are to imagine what would follow if it were a sufficient condition for a man to be justified in holding a belief P that 'he holds an infinite (possibly circular) sequence of beliefs $Q_1, Q_2, \ldots$ such that P is supported by some of the beliefs in the sequence and each belief in turn is supported by later beliefs in the sequence'. If this condition were sufficient, a man could be justified in believing anything. 'For example, given an arbitary belief P, S would be justified in holding it if he also happened to believe each of

$Q_1, Q_1 \supset P$;
$Q_2, Q_2 \supset Q_1, Q_2 \supset (Q_1 \supset P)$;
$Q_3, Q_3 \supset Q_2, Q_3 \supset (Q_2 \supset Q_1), Q_3 \supset (Q_2 \supset (Q_1 \supset P))$;
$Q_4, Q_4 \supset Q_3, Q_4 \supset (Q_3 \supset Q_2), Q_4 \supset (Q_3 \supset (Q_2 \supset Q_1)), Q_4 \supset (Q_3 \supset (Q_2 \supset (Q_1 \supset P)))$;
$Q_5 \ldots$

In this sequence of beliefs, each belief is supported by beliefs later in the sequence, but the beliefs are nowhere tied down in any way to the evidence of S's senses. As long as a person's beliefs form such a coherent set, he could hold any beliefs at all regarding the colours, shapes, sizes, etc., of things, regardless of how they look or feel to him.' He could be justified in believing that all of his senses mislead him in a systematic way. But 'it is impossible for a person to be justified in believing that *all* of his senses systematically mislead him *all* of the time'. Hence 'there are some (epistemologically basic) beliefs, which we can be justified in holding without being able to justify them on the basis of other beliefs' (J. Pollock, *Knowledge and Justification*, Princeton 1974, pp. 27–8).

If we forget for the moment that q is sufficient evidence for p only if q is true and assume that '$q \supset p$' is a sufficient condition for 'q is sufficient

evidence for p', then the considerations which Pollock adduces would remind us of the falsity of

> (A) It is a sufficient condition for N's being justified in accepting a proposition p that he believes each member of an infinite series of propositions such that some proposition in the series is sufficient evidence for p, and such that, for each proposition i in the series some later proposition is sufficient evidence for i.

But what I am trying to find is an argument for the falsity of

> (B) It is a sufficient condition for N's being justified in accepting p that he is justified in believing each member of an infinite series of propositions, such that his being justified in believing one member is sufficient for his being justified in believing p, and such that, for each proposition i in the series, his being justified in believing a later proposition is a sufficient condition for his being justified in believing i.

Taken as an argument for (B)'s falsity, Pollock's reasoning would be question-begging. It could be put in the form of a reductio, as follows. Suppose (B) is true, and suppose also, what is obviously possible, that N does in fact believe that his senses are totally unreliable, and is also justified in believing each member of an infinite series of propositions, such that his being justified in believing one member is sufficient for his being justified in accepting that his senses are totally unreliable, and such that for each proposition i in the series his being justified in believing a later proposition is a sufficient condition for his being justified in believing i. Then N would be justified in believing that his senses are totally unreliable. But this is obviously absurd. So (B) is false. The trouble is that any normally intelligent person who was uncertain about whether (B) was true or false would be equally uncertain about whether to accept one of the premises necessary for deriving the absurdity. If you think it absurd to suppose that N is justified in believing that his senses are totally unreliable, then this is because you think it obvious that this proposition about N's senses is false and take it as an elementary logical implication of 'N is justified in believing p' that p is probably true. But in this case you will be equally disinclined to accept the premiss that N is justified in believing each member of an infinite sequence of propositions, such that his being justified in believing one of them is sufficient for his being justified in believing that his senses are totally unreliable. A regress stretching back from an obviously false belief, you might say, obviously can't be a regress of *justification*. If you *weren't* equally disinclined to accept this premiss, it would be because you were mixing it up with the quite different proposition that there can be an infinite series of propositions, such that N believed each member of it, and such

that some propositions in the series are sufficient evidence for the proposition that N's senses are totally unreliable, and for each proposition i in the series, some later proposition is sufficient evidence for it. There is no impossibility here, if we assume that q is sufficient evidence for p when it materially implies it. For on this assumption a man can believe two propositions, one of which is sufficient evidence for the other, even though it is more likely than not that both propositions are false.

Secondly, there is an argument which turns on the alleged impossibility of performing an infinite number of tasks in a finite time. I am assuming that becoming indirectly justified in believing p requires an activity of investigation. This makes it natural to suppose that a regress of justification must terminate in directly justified believing, for otherwise it would be impossible for a man to become indirectly justified in believing any proposition in a finite time without completing an infinite number of separate investigatory tasks in a finite time. But if the object is to show that there is an actual contradiction in the notion of an infinite regress of justification, then this reasoning is not conclusive. Even if we suppose it necessarily true that a man indirectly justified in believing p *had* only a finite time in which to complete his justificatory tasks, the conclusion still doesn't follow. Suppose it took the man less time to consider whether p_2 was sufficient evidence for p_1 than to consider whether p_1 was sufficient evidence for p, less time to consider whether p_3 was sufficient evidence for p_2 than to consider whether p_2 was sufficient evidence for p_1 and so on ad infinitum. Unless there was some minimum time it took him to consider whether any proposition was sufficient evidence for another, then he could consider each of an infinite number of evidential relations in a finite time, just as he could, in a finite time, perform the infinite number of actions which constitute traversing the infinite number of smaller and smaller intervals involved in running a mile (cf. C. S. Peirce, 'Questions concerning certain faculties claimed for man', *Philosophical Papers* V, pp. 154–5). Why should we suppose it necessarily true that there is any such minimum time?

III

If indirect justification is not heteronomous, then you can be justified in believing p, but not directly, and not because you are directly justified in believing some other proposition which you take to be sufficient evidence for p, nor yet because you are directly justified in believing a third proposition which you take to be sufficient evidence for the proposition which you take to be sufficient evidence for p, nor yet . . ., and so on ad infinitum. I'm going to say that anyone who is justified in believing p under these conditions is α-justified in believing p, and cast

my argument for the heteronomy of indirect justification in the form of an argument for the impossibility of α-justification.

Consider the following train of thought. Either we can make sense of the idea of a denumerably infinite conjunction or not. If we can't, then we can't make sense of the idea of a potentially infinite conjunction either. 'A strict finitist would deny the (constructible) existence of infinitely proceeding sequences in very much the same manner as the intuitionists deny the existence of actual infinities ... We can imagine a process of stroke added to stroke, up to a point: but there comes a point after which perception and intuition no longer keep pace' (S. Körner, *The Philosophy of Mathematics*, London 1960, p. 148). So either (i) it is prima facie intelligible to talk of the infinite conjunction of all those propositions which a man will be α-justified in believing if he is α-justified in believing any proposition, or (ii) the notion of α-justification is immediately unintelligible. But even if we accept alternative (i) α-justification is unintelligible in the end. If C is the conjunction of all the propositions which the man is α-justified in believing, he will be justified in believing C itself. But this means that there is a proposition q, not identical to C, which he is justified in taking to be sufficient evidence for C and which he is α-justified in believing. But he can't be justified in taking q to be sufficient evidence for C if q is an actual *conjunct* of C. And yet if he is α-justified in believing q it *must* be such a conjunct.

The reasoning from alternative (i) is obviously grossly defective as it stands. Why should it follow from your being justified in believing p and justified in believing q and . . ., that you are justified in believing (p & q & . . .,)? And even if this did follow, we couldn't even initially suppose that the man was α-justified in believing C without absurdly supposing that C was a member of itself. But here is a revised version of the non-finitist sector of the argument, designed to circumvent these objections.

(1) N is α-justified in believing some proposition.

(2) If N is α-justified in believing a conjunction X, then there is some set of non-conjunctive propositions whose conjunction is identical to X and each member of which N is α-justified in believing.

(3) If N is α-justified in believing some proposition, then there is some non-conjunctive proposition which he is α-justified in believing (from (2)).

(4) Let us say that a conjunction is someone's α-conjunction if it is the non-repetitive conjunction of those non-conjunctive propositions each of which he is α-justified in believing.

(5) If N is α-justified in accepting some proposition, then he has an α-conjunction C, which it is logically possible for him to be α-justified in believing.

(6)   If N were α-justified in believing C, it would be because there was a
      proposition q, non-identical to C, which he was α-justified in
      believing and which he was justified in believing to be sufficient
      evidence for C.

(7) q is either a conjunctive or a non-conjunctive proposition.

(8)   If q is non-conjunctive then it will be a conjunct of C (from (4) and
      (6)).

(9)   If q is conjunctive, and N is α-justified in believing q then there is
      some set of non-conjunctive propositions whose conjunction is
      identical to q, and each member of which N is α-justified in believ-
      ing (from (2)).

(10) If q is conjunctive, then q is identical to a conjunction of some of C's
      conjuncts.

(11) N cannot be indirectly justified in believing a conjunction X
      because he is justified in believing a conjunct of it, or because he is
      justified in believing a proposition identical to a conjunction of
      some of X's conjuncts, unless he is justified in believing that this
      conjunct, or conjunction of X's conjuncts, is sufficient evidence for
      X.

(12) N cannot be justified in believing that this conjunct or conjunction
      of X's conjuncts is sufficient evidence for X.

(13) N cannot be α-justified in accepting C (from (6), (7) ,(8), (10), (11),
      (12)).

(14) N cannot be α-justified in believing any proposition (from (5) and
      (13)).

I'll now say something about the more questionable premisses of
this argument. Some of them turn out to be truths about quite ordinary
concepts which we obviously want to apply. Others hold only because of
what I'll build into, or have already built into, the definition of indirect
justification; whether or not these stipulations trivialize my conclusion is
a question I postpone until section IV. One vital premiss I must simply
leave to stand or fall with the feasibility of a certain way out of the lottery
paradox.

I shall take it as necessary and sufficient for two propositions to be
identical that it is logically impossible for someone to believe one of
them without believing the other. On this condition logically equivalent
propositions will not necessarily be identical: it is logically possible to
believe p without believing (p & q), where 'q' stands for a logically
necessary proposition. But some logically equivalent propositions will
nevertheless be identical. Thus ((p & q) & (r & s)) will be identical to
(p & q & r & s), and in general

(C)   A conjunction of conjunctions is identical to some conjunction of
      non-conjunctive propositions with which it is logically equivalent.

Now, why should we accept (2)? It seems reasonable to assume that although

(D) 'N believes p and N believes q and . . .' does not entail 'N believes $(p \ \& \ q \ \& \ . . .)$'

nevertheless

(E) 'N believes $(p \ \& \ q \ \& \ . . .)$' entails '$N$ believes p and N believes q and. . . .'

correspondingly

(F) 'N is justified in believing $(p \ \& \ q \ \& \ . . .)$' entails '$N$ is justified in believing p and justified in believing q and. . . .'

It seems hard to deny (2) if you accept both (F) and my criterion of propositional identity.

Why should we accept (5)? Given (D), we must admit that

(G) 'N is justified in believing p and N is justified in believing q and . . .' does not entail 'N is justified in believing $(p \ \& \ q \ \& \ . . .)$'.

But it may still be true that

(H) 'N is justified in believing p and N is justified in believing q and . . .' entails 'it is logically possible for N to be justified in believing $(p \ \& \ q \ \& \ . . .)$'

just as it seems true that

(J) 'p entails q and N is justified in believing p' entails 'it is logically possible for N to be justified in believing q'.

If (H) were true, it would be reasonable to assume also that

(K) 'N is α-justified in believing p and N is α-justified in believing q and . . .' entails 'it is logically possible for N to be α-justified in believing $(p \ \& \ q \ \& \ . . .)$'.

And if (K) is true it is hard to deny (5).

(H), however, collides with the lottery paradox. Consider the conjunction of (a) 'one of the faces 1, 2, 3 and 4 will turn up on the next throw', (b) 'one of the faces 3, 4, 5 and 6 will turn up on the next throw', and (c) 'one of the faces 1, 2, 5 and 6 will turn up on the next throw'. A man might be justified in accepting each of (a), (b) and (c). But it is not logically possible for him to be justified in accepting their conjunction, because their conjunction is self-contradictory, and this seems not to square with the requirement that if you are justified in believing p then p is more likely true than false. The topic is too vast to discuss in this

paper. All I can do is mention one possible way out. This is to claim that it only seems obvious that a man might be justified in accepting each of (a), (b) and (c) because we implicitly treat (a) as short for 'it is more probable that one of the faces 1, 2, 3 and 4 will turn up on the next throw than that 5 or 6 will'. There is nothing self-contradictory about the set of propositions which consists of this full version of (a) and similarly full versions of (b) and (c) (see Richard Swinburne, *Introduction to Confirmation Theory*, London 1973, p. 187).

What about (6)? I began by making it a necessary condition for you to be indirectly justified in believing p not just that there is some other proposition q which you believe and believe to be sufficient evidence for p, but also that you are justified in believing q. If (6) is to hold yet more must be built into the definition of indirect justification. We must make it necessary not only that you are justified in believing this proposition q but also that you are justified in believing that q is sufficient evidence for p. With indirect justification defined in this way, (11) holds too.

Finally, why (12)? I've been taking it that if you are justified in believing p then p is at least probably true. But it seems to me that if 'sufficient evidence' is taken in its ordinary sense, then if N did believe that a conjunct of X or a conjunction of some of X's conjuncts was sufficient evidence for X he would believe something which was necessarily false. I said at the beginning that q isn't sufficient evidence for p if the argument 'q therefore p' is question-begging, and that this is what makes it incorrect to say that a proposition is sufficient evidence for itself. Now, it is true enough that an argument to X from a conjunct of X or from a conjunction of some of its conjuncts need not be question-begging in the sense I intended. A normally intelligent man who was doubtful about the truth of 'Swan$_1$ is white and swan$_2$ is white and swan$_3$ is white and swan$_4$ is white' wouldn't necessarily be equally doubtful about the truth of 'Swan$_1$ is white and swan$_2$ is white and swan$_3$ is white'. However, it does also seem to be part of the ordinary concept of sufficient evidence that

(L) If a proposition is sufficient evidence for a conjunction then it is sufficient evidence for each conjunct of that conjunction.

And if you conjoin (L) with the condition excluding question-begging-ness then my conclusion follows. Can a conjunct of X be sufficient evidence for X? If so, then (L) entails that this conjunct will be sufficient evidence for itself. Can a conjunction Z of some of X's conjuncts be sufficient evidence for X? By (L) Z would have to be sufficient evidence for each of X's conjuncts. But Z has no conjuncts which are not also conjuncts of X. So Z will have to be sufficient evidence for each of its own conjuncts. And this is just as absurd as the conclusion that Z is

sufficient evidence for itself. If p is a conjunct of q, the argument 'q therefore p' is just as question-begging as the argument 'q therefore q'.

IV

As I've defined it, then, indirect justification is heteronomous. If there is any proposition which you are indirectly justified in believing, then there is some proposition which you are directly justified in believing. But in order to make the argument work I've had to build rather a lot into the definition of indirect justification. (6) wouldn't have held if I hadn't laid it down that when a man is indirectly justified in believing p by virtue of adducing a proposition q as sufficient evidence for p, he is also justified in believing q itself. But mightn't we have circumstances in which, as Austin put it, q doesn't '*need* to be verified' (J. L. Austin, *Sense and Sensibilia*, Oxford 1962, p. 117)? And then again, do we really want to insist, as I had to in order to make (11) hold, that when a man is indirectly justified in believing p by virtue of adducing q as sufficient evidence for p, he is justified in believing that q is sufficient evidence for p. Braithwaite's *Scientific Explanation* contains vigorous arguments to the effect that in order to be reasonable in believing a proposition which we have inferred from something we believe, we don't need to be reasonable in accepting the effectiveness of the policy in accordance with which we have made the inference (R. B. Braithwaite, *Scientific Explanation*, Cambridge 1953, Ch. 8). If 'direct justification' covers all justification which isn't on my definition indirect then doesn't it cover too much for there to be any interest in the heteronomy of what it holds in thrall?

To my mind the interest of the heteronomy thesis lies in the rather plausible form of scepticism which it suggests or at any rate facilitates. What I've shown is that if there is any proposition p which you are justified in believing only because you are justified in believing some other proposition q and justified in believing that q is sufficient evidence for p, then (i) there is some proposition r which you are justified in believing even though either (a) there is no proposition which you take as sufficient evidence for r, or (b) there is no proposition which you both take as sufficient evidence for r and are justified in believing, or (c) there is no proposition which you are both justified in believing and justified in believing to be sufficient evidence for r. Furthermore, (ii) r must be identical either to q, or to a proposition which you are justified in believing to be sufficient evidence for q, or to a proposition which you are justified in believing and justified in believing to be sufficient evidence for a proposition which you are justified in believing and justified in believing to be sufficient evidence for q, or . . ., and so on. A sceptic can claim with some plausibility that on a generic concept of justification which we actually want to apply, there are rather few values of r which you can be justified in believing when either (a) or (b) or (c) is satisfied,

and hence, by (ii), rather few propositions which you can be either directly
or indirectly justified in believing in this generic sense. In broad outline,
the sceptical argument would go like this. We do indeed want a justified
believing which requires some conscious exercise of reason on the be-
liever's part. And it seems that you can't be in this sense justified in
believing r when (a) is true unless r is one of those rare propositions
which you can simply 'see' to be true. There is no room here for 'justi-
fication by experience': having an experience need involve no conscious
exercise of reason. You can't be justified in believing r when (b) is true
unless either (a) is true or the proposition which you adduce as evidence
for r is one which you somehow don't need to be justified in believing.
But if what we need depends on what it is *good* for us to want, then it is
not easy to see how exactly the non-existence of a need for justification is
going to be established. It certainly won't follow simply from the
unsatisfiability of the corresponding want.

Finally, you can't be justified in believing r when (c) is true unless
either (a) or (b) is true, or you somehow don't want or need to be
justified in accepting the effectiveness of the policies in accordance with
which our inferences are made. But what goes for our wants and needs
in relation to premisses goes also for our wants and needs in relation
to inferential policies.

University of East Anglia   © N. M. L. NATHAN 1977

ON SQUARING SOME CIRCLES· OF LCGIC

By James J. Strom

PERENNIAL though perhaps superficial doubts abcut the validity of logic are commonly encountered, among othe⁻ places, in introductory courses. Recently ('On the Logic of the Circularity of Logic', *Mind*, January, 1975) G. B. Keene has, in attempting to assess such doubts, suggested two arguments which together with a completeness theorem for the predicate calculus produce, he thinks, first an apparent paradox and then this resolution: 'The very enterprise cf attempting to formulate an argument to prove that no set of rules cf inference can be shown to generate only valid arguments, carries its own unfeasibility with it.'

Each of his two arguments contains an invalidating oversight; in straightening them out we should be able to understand better the role we assign to formal logical systems.

In his first argument Keene uses '$G(x)$' to abbrev₋ate 'The set of rules x generates only valid arguments' and 'A' to designate an argument the conclusion of which is that the set of rules R 'generates only valid arguments (i.e., $G(R)$)'. He observes

> (2) that A succeeds in establishing $G(R)$, presupposes $G(R)$ (in so far as A uses R),

and concludes, after showing that A is question-beggir₋g, that

> (5) any argument purporting to prove that an⁻ set of rules of inference generates only valid arguments, begs the question, [and]
> (6) no set R of rules of inference can be shown to generate only valid arguments.

Keene believes that this first argument, which he calls ϕ, results in an apparent paradox when combined with a completeness theorem. (I construe Keene as meaning a *soundness* theorem, since (6) would conflict with proofs of soundness, but not of completeness.)

The second of Keene's arguments must refute or at least undercut ϕ. If it does not, we may have to accept ϕ's perfectly general conclusion that no argument can succeed in rationally justifying any set of rules, particularly including those of any logic. Keene's strategy is to show that ϕ self-referentially implies its own inadequacy and thus that ϕ itself is self-defeating. One appealing way to do this is to take ϕ to demonstrate not merely that there are no proofs for the soundness cf rules (its actual conclusion), but that no rules are sound. Then ϕ itself would not be valid.

127

Surprisingly, Keene takes φ as showing just this. His second argument opens with an imperspicuous translation of (6) as $\sim\exists x G(x)$. This would approach adequacy only if rules which cannot be proved sound are not sound; i.e., do not generate valid arguments only. The details of this translation do not formally enter his argument, but at [4] of his second argument Keene makes just the kind of shift his translation suggests:

> [4] any argument which uses rules not known to generate only valid arguments, cannot establish its conclusion.

From this Keene concludes that φ fails to establish its conclusion, and goes on to suggest the resolution I mentioned above.

Now this, particularly [4], appears to be a mistake. But perhaps I have misread Keene's vocabulary. Keene's 'prove' and 'show' are synonymous, and 'not known' is understood as entailed by 'not provable' in the case of the soundness of rules: so much is textually apparent. However, I also take *establishing* to be the same thing as *proving*, but perhaps 'establish' means more than 'prove', saving the truth of [4]. Although Keene apparently uses these words in (2) and (3) as synonyms, 'establish' may, in [4], mean *prove and have rules of proof known to be sound* or the like. But while this stronger meaning would make [4] true, it would also make the second argument irrelevant to φ. For while now φ may be said not to *establish* its conclusion, it may still validly *prove* its conclusion— and if φ is a proof, logic is still in trouble.

In [4], then, either 'establish' means 'prove' and [4] is relevant to φ but assumes that rules not proved to be sound are not sound—or 'establish' has a stronger meaning including that the rules of proof are known to be sound, but in this case [4] has no bearing on φ unless we believe that an argument which does not establish its conclusion also does not prove it. So no matter which way we interpret [4], the key to understanding Keene's argument is his peculiar translation of (6), which identifies, or confuses, soundness with demonstrated soundness. Unless we can accept such an identification, which itself would render logic an epistemological morass, we shall think that Keene has not defeated φ. This is the mistake in Keene's second argument.

However, there is a mistake in his first argument too. We need not fear that all arguments for rules are question-begging, unless we are willing to make a seemingly implausible assumption: we need not assume, in (2), that A uses R. The argument φ is formulated to cover any argument and any set of rules; thus (2) succumbs to innumerable counterexamples. For example, the reasoning establishing derived rules, such as De Morgan's, within a formal calculus, need not presuppose those derived rules. Similarly, rules can be supported from outside, as when informal arguments are used to set up a formal system itself. Thus the

broad (5) and (6) do not follow without the substantive assumption Keene includes in (2).

Clearly φ would be more convincing if it were formulated to apply to the all-encompassing logic suggested by Keene's introductory comments. How do we argue for or against, or revise, the totality of our method of reasoning? If M is the set of all underivative principles of inference and axioms which generate only valid arguments, any argument for M had better be question-begging—or else it is invalid. If we were to argue successfully for M on the basis of B, a set of principles outside M, M would not satisfy the definition above; it would in effect have been replaced by B. However, it may be a mistake to conclude from this that M has no foundation or is founded in 'the irrational', though its foundation will not be a standard M-sanctioned argument for M as a whole.

Showing in what respects an argument like φ would be correct hardly 'carries its unfeasibility with it' without also carrying Keene's apparent belief that if you can't prove a rule it can't be sound. However, showing this does not clash paradoxically with soundness proofs which purport to cover an area smaller than or derivative from M.

Keene began his discussion with the implicit confidence that the principles of inference of elementary logic can 'provide a formal device by which the validity of any argument (which *is* valid) can be established'. But if there is a lesson in these reasonings it is the error of thinking that a formal calculus for which we could offer a soundness proof could be identical with what I call M. If it were, any soundness theorem for that calculus would in fact be question-begging; on the other hand, if we are willing to argue the soundness of a calculus we show that we do not believe it is the whole ball of wax.

Yale University

IMPLICIT BARGAINING AND MORAL BELIEFS

By CHRISTOPHER NEW

IN his recent paper, 'Moral Relativism Defended' (*Philosophical Review*, January 1975, pp. 3–22), Professor Gilbert Harman puts forward a hypothesis which he thinks explains otherwise puzzling features of our moral beliefs better than any other theory known to him. The hypothesis is that some at least of our moral principles represent a compromise reached by means of 'implicit bargaining between people of varying powers and resources' (pp. 11–15). And the puzzling features of our moral views which this hypothesis is supposed to explain are that we attach a greater weight to the duty not to harm others than we do to the duty to help others and that we consider every person has an inalienable right to self-defence and self-preservation. I shall show that the hypothesis in no way explains these features of our moral views. (It might be doubted whether there is general agreement that it is *always* worse to harm others than not to help them or that *everyone* has an inalienable right to self-defence; but I shall let that pass. My aim is to show that if we do believe we have these duties and rights, Harman's hypothesis fails to explain why.)

Harman does not tell us much about what implicit bargaining is, but apparently it occurs when members of a group form intentions to act in their own interests and then discover that other members of the group have formed different intentions in *their* own interests. The intentions conflict and it is then that 'implicit bargaining' takes place, which issues in 'some sort of compromise'. The members of the group eventually reach an 'agreement in intentions'—each member of the group forms the intention to act according to certain principles on condition that the other members similarly intend (p. 13; it is not clear whether this is supposed to happen in the biography of every member of a group or only in that of some; but, again, I shall let that question pass, since its answer would not affect the point I wish to make).

How does Harman think this hypothesis will explain these otherwise puzzling features of our moral beliefs? Let us take the case of our differently weighted duties first. Harman notes that we would think it wrong of a doctor to save five of his patients who would otherwise die by cutting up a sixth patient and distributing his healthy organs where needed to the others, although we also think a doctor has a duty to save as many of his patients as he can. We would think it wrong because we believe the doctor has a stronger duty not to harm any of his patients (p. 12). This appears hard to explain, Harman argues, unless we suggest that our beliefs are the result of implicit bargaining among people of varying powers and resources. For everyone, rich, poor, strong or weak,

would benefit from and therefore accept a principle that all should avoid harming each other. But only the poor and weak stand to benefit from a principle that everyone should help others who are in need. Consequently, he concludes, the rich and strong would be unwilling to accept the second principle and a compromise would be likely to develop, in which the duty to help others was assigned less weight than the duty to avoid harming them.

The second case—of the right to self-defence and self-preservation—is closely related, one would think, to the first one. For such a right seems to be merely the correlative of the duty to avoid harming others. If the duty to avoid harming others could gain universal approval, then the right to defend and preserve oneself would surely gain it too. So, at least, one would expect Harman to argue, but, curiously, he does not. He appears to argue instead that, since it is impossible (except in very special circumstances) rationally to form the *intention* not to try to save your life—for you know that when the time comes you *will* try to save it—it is impossible for you to intend to keep an *agreement* not to do so—that is, to intend not to do so provided others similarly intend (pp. 14–15).

This argument is an extraordinary one. It is of course perfectly valid, but it by no means shows that the belief in our right to self-defence and self-preservation can be explained by the hypothesis that it derives from a process of bargaining leading to an implicit agreement. What it shows is that since we could not *intend*, we could not *agree* in intending not to do what we know we will do. And that does not show what Harman's argument seeks to establish, that we must therefore have agreed in intending to *do* what we know we will do. (Similarly we cannot agree in intending not to feel *our* pains, for any pain we feel is ours. But that does not show that we must have agreed in intending to feel only our pains.) His point is quite irrelevant to the implicit bargaining and agreement thesis. That hypothesis is better served by arguing, as I have suggested, that the right to self-defence and self-preservation is a correlative of the duty not to harm others, and I shall accordingly treat it as such in what follows. The features of our moral views which Harman hopes to explain are thus not really two different features but two aspects of the same one.

It has been argued against Harman's explanation of one of these aspects—the belief that we have a greater duty to avoid harm than to provide help—that there is an alternative and equally satisfactory explanation available (Robert Coburn, *Philosophical Review*, January 1976, pp. 87–93). This alternative explanation is that our behavioural dispositions have evolved, by a process of natural selection involving the genetic basis of those dispositions, in such a way that a strong principle of harm avoidance and a weak principle of mutual aid have become prevalent in our society.

It is not at all clear, though, that this 'alternative' explanation is incompatible with Harman's. For Harman's hypothesis may be concerned with the question 'What considerations lead (have led) people to adopt these principles?', whilst the purported alternative is concerned with the question 'How has it come about that people adopting these principles have survived in our society and even flourished?' And it is surely conceivable *both* that people adopt (have adopted) these principles after a process of implicit bargaining in the way Harman suggests *and* that natural selection has favoured those genes or gene clusters which underly the adoption of, and behaviour conforming to, them. On this view, the two explanations may be complementary, rather than competitors.

However, I am not a confident interpreter of Harman's intentions. Perhaps he wants to say not merely that the best explanation of why we chose these differently weighted principles is that there has been an antecedent process of implicit bargaining, but also that that process and the resultant behaviour has no underlying physical or genetic basis. In that case, of course, his hypothesis, if it were tenable in itself, would be a genuine competitor with the evolutionary one. But fortunately there is no need to go further into the question of Harman's intentions. Whatever they are, the hypothesis itself is not a competitor with the evolutionary or any other hypothesis. For non-starters cannot compete and Harman's hypothesis is a non-starter, as the following considerations show.

Let us examine first the hypothesis's claim to explain the existence of the strong duty to avoid causing harm, together with which, I have suggested, goes the correlative right to self-defence and self-preservation. Why do we have a strong principle of harm avoidance? According to Harman's hypothesis, we must suppose that there are people of varying powers and resources bargaining from the point of view of pure self-interest. (That they not only have, but know they have, varying powers and resources is a crucial feature of Harman's theory, which distinguishes it sharply from a Rawls-type theory in which rational self-interested agents are represented as choosing in ignorance of their powers and resources.) 'The rich, the poor, the strong and the weak', he says, 'would all benefit if all were to try to avoid harming one another. So everyone could agree to that arrangement' (p. 12). Hence a strong duty of harm avoidance (and—as I have amended the hypothesis—a correlative right of self-defence and self-preservation) comes about.

But *would* they all benefit? Suppose we are both members of a group and you are very strong but poor and I am very weak but rich. What is there for you in an arrangement whereby I do not intend to harm you so long as you do not intend to harm me? Obviously, if you are strong, you are not likely to suffer from my intention to harm you and I, indeed, being rich, will have no motive, independently of your actions, for

forming any such intention. But you *do* stand to gain by, and *have* a motive for, intending to harm me, because you are not likely to come off worse and you will, if successful, obtain my riches. So there is no reason at all for you to think that you would benefit from such an arrangement. It is simply false, from the point of view of sheer self-interest, that *all* would benefit from the adoption of such a principle. Consequently, of course, it is also false that *all* would benefit from the granting to all of the correlative right to self-defence and self-preservation.

Now let us examine the case of the weaker duty to help those in need. 'The rich and the strong would not benefit from an arrangement whereby everyone would do as much as possible to help those in need,' Harman argues, 'the poor and weak would get all of the benefit of the latter arrangement.' Consequently the rich and the strong 'would be reluctant to agree to a strong principle of mutual aid. A compromise would be likely and a weaker principle would probably be accepted' (pp. 12–13). This, he thinks, explains why we believe the duty to help others is weaker than that to avoid harming them. The belief derives from the fact that some members of the group would not see anything for them in a principle of mutual help. Well, of course Harman is right if he thinks that the irch and strong would get nothing out of such an arrangement. But that merely shows that the two supposedly different cases are really the same— it is true in both cases that the arrangement would not benefit everyone. In the case of a proposed principle of harm avoidance *and also* in the case of a proposed principle of helping those in need, some members of the group would stand to gain nothing at all from the arrangement. And since the cases are the same in that respect, the implicit bargaining hypothesis cannot be used to explain why we believe (so far as we do) the duty to avoid causing harm is stronger than the duty to provide help.

Harman's hypothesis, then, does not explain at all, let alone better than anything else, why we hold these beliefs.

University of Hong Kong   © CHRISTOPHER NEW 1977

ARTIFACTS, NATURAL OBJECTS, AND WORKS OF ART

By DANIEL DEVEREUX

WHAT is required for a natural object to become an artifact? Is it enough, as George Dickie suggests, that the object be treated as a kind of artifact?[1] I want to argue that it is not enough—that for a natural object to be 'artifactualized' it must at least undergo a change in the thing itself, and not just in the relationships which it has to other things. The relevance of this question to the issue of the definability of art will emerge in the course of the discussion.

Dickie argues against Morris Weitz's claim that a natural object, like a piece of driftwood lying on the beach, could be a work of art. If someone were to call it a work of art, he would, according to Dickie, be using the expression in its evaluative and not its descriptive sense; i.e., he would be praising the object rather than classifying it.[2] But in discussing the question whether a work of art must be an artifact, and the more general question whether the expression 'work of art' is definable, we are interested in the descriptive or classificatory use of this expression and not in its evaluative use. Thus the fact that someone might call a piece of driftwood on the shore a work of art does not show that a natural object could be a work of art in the relevant sense.

But Dickie goes on to say that a piece of driftwood could *become* a work of art if, for example, someone carried it home and hung it on a wall.[3] This does not upset his general claim that a natural object could not be a work of art, for he argues that the very process which makes the driftwood a work of art also makes it an artifact. In other words, putting a piece of driftwood on display, in the way in which one might put a painting or other work of art on display, transforms it from a natural object into an artifact.

One may feel an initial scepticism about the idea that a natural object can become an artifact simply by being put on display in a certain way. That there is a solid basis for this feeling can be brought out by the following examples. (1) Suppose I am out in the woods looking for a fern to take home with me. I find a rather attractive one and transplant it to a pot. When I get home I hang the pot in the living room in a place which will show off the plant to best advantage. I wish to display it for basically the same reason I would a beautiful picture. On Dickie's view we must

---

[1] *Art and the Aesthetic* (Cornell University Press; Ithaca, N. Y., 1974), pp. 44–45. Cf. also 'Defining Art' *American Philosophical Quarterly* (1969), pp. 253–256.

[2] 'Defining Art', p. 253. In *Art and the Aesthetic* Dickie modifies his objection to Weitz by introducing a third sense of 'work of art' (pp. 24–26). Since this issue is not relevant to my argument, I have stuck with the simpler formulation.

[3] *Art and the Aesthetic*, p. 44 ('Defining Art', p. 255).

say that the fern is no longer a natural object, that it has been trans-formed into an artifact. (2) Presumably it is not essential to the process of transformation that the object be put on display inside a building. Suppose I buy a tree from a nursery and carefully choose a place in my front yard to plant it. My reasons for choosing this particular tree and this particular place for it are chiefly aesthetic. Does the tree, once planted, become an artifact? Does it cease to be a natural object?

As I think these examples show, for a natural object to become an artifact it is not enough that it be displayed in a certain way. As a general rule, I suggest that a natural object cannot become an artifact without undergoing some internal change. By 'internal change' I mean a change in the thing itself and not just in its relationship to other things. Changes in size, shape, or colour, would count as internal changes; a change in physical location would not. I am not suggesting that if I produce an internal change in a natural object it thereby becomes an artifact; pruning a tree does not make it an artifact. My claim is that the process of trans-forming a natural object into an artifact must *at least* involve some inter-nal change in the object.

This claim is open to question. Suppose I find a large piece of drift-wood on the beach which has the general shape of a chair and is very comfortable to sit on. I take it home, and without doing anything to it I start using it as a chair. After a while I no longer think of it as a piece of driftwood which I am using as a chair, but simply as one of my chairs. It might be argued that since a chair is a kind of artifact, this piece of driftwood has become an artifact without undergoing any internal change.

I would grant that it is natural and appropriate to speak of the driftwood in this example as a chair. But it is not necessary to accept the claim that a chair must be an artifact. Paperweights and bookends which are not artifacts are more common than chairs which are not, but this makes no difference to the argument.

If it is true that a natural object cannot become an artifact without undergoing an internal change, then the driftwood which is hung on the wall is not an artifact. Of course the driftwood could become *part* of an artifact without any internal change. For example, I might use a small piece of driftwood in putting together a collage.

The results of the discussion so far raise a question about Dickie's general claim that being an artifact is a necessary condition of being a work of art. For if the driftwood hanging on the wall is a work of art, as he claims, then it seems that Weitz was right in arguing that a work of art need not be an artifact. But, as far as I can see, there is nothing which compels us to regard the driftwood as a work of art. We enjoy natural beauty as well as art, and we sometimes treat natural objects in very much the same way that we treat works of art. The fern hanging in someone's

living room may have the same function as the painting on his wall. But that is not a good reason for considering the fern a work of art. And since the driftwood hanging on the wall need not be classified as a work of art,[1] it does not undermine Dickie's general thesis that artifactuality is a necessary condition of being a work of art.

[1] One could perhaps consider the driftwood together with the wall as a composition and a work of art. But in this case the driftwood would only be a part of a work of art; it would not itself *be* a work of art.

University of Virginia   © DANIEL DEVEREUX 1977

MORE ON QUINE'S REASONS FOR INDETERMINACY OF TRANSLATION

By ROBERT KIRK

QUINE has argued that the underdetermination of theories brings with it indeterminacy of translation of theories: translation of a theory is indeterminate at least to the extent that the theory is underdetermined by the totality of true observation sentences, so that where *A* and *B* are physical theories both compatible with the data, it might be possible for us to 'adopt *A* for ourselves and still remain free to translate the foreigner either as believing *A* or as believing *B*' ('On the Reasons for Indeterminacy of Translation', *Journal of Philosophy*, 67 (1970), 178–83). I have tried to show that Quine's argument fails ('Underdetermination of Theory and Indeterminacy of Translation', ANALYSIS, 33 (1973), 195–201). M. C. Bradley and Peter Smith have each raised counter-objections (M. C. Bradley, 'Kirk on Indeterminacy of Translation', ANALYSIS, 36 (1975), 18–22; Peter Smith, 'Kirk on Quine's Reasons for Indeterminacy of Translation', *Philosophical Studies*, 27 (1975), 427–31). The extraordinary interest of Quine's thesis of the indeterminacy of translation and the dearth of good arguments either for or against it are, I think, good reasons for taking the discussion further.

I

First, an objection of Bradley's to my whole strategy. I maintained that if Quine's argument works at all, it should work even if we assume that the underdetermination of theory sets in only at a certain level of theoreticity, and that translation of sentences below that level is determinate. I went on to argue that on this assumption there will also be

determinate translation above that level of theoreticity, even if the theories to be translated are underdetermined. To Bradley 'it seems most doubtful whether this is a permissible form of argument against Quine' (*op. cit.*, p.21). Bradley's reason is that Quine argues elsewhere: (i) (from the alleged underdetermination of parsing) to the effect that indeterminacy of translation cannot be confined to the remote theoretical terms of the language; and (ii) that 'given indeterminacy of translation of terms *somewhere*, the translation of quantifiers is rendered indeterminate *everywhere*' (*op. cit.*, pp. 21f.). Now it seems to me that both (i) and (ii) are quite compatible with the assumptions and strategy of my argument, which is after all intended to show only that Quine's argument in 'On the Reasons . . .' does not work, not that its conclusion is false or that his other arguments for indeterminacy are also unsound. However, I evidently failed to make clear the role in my argument of the assumption that translation is determinate up to a certain point. I think Bradley's misgivings will be seen to be unfounded if the following sketch succeeds in clarifying the role of this assumption.

Let P = *underdetermination sets in at that level of theoreticity at which theory X is introduced.*

Q = *translation of theory X is indeterminate.*

R = *translation is determinate up to that level of theoreticity at which theory X is introduced.*

(This is the assumption in question.)

Quine argues that

(1)   P entails Q.

Against this I argue that

(2)   if R, then P does not entail Q.

Bradley points out that Quine has (i) an argument which purports to establish that

(3)   translation is indeterminate at all levels,

which entails the falsity of R; and (ii) an argument to the effect that

(4)   Q entails not–R.

I now reply that if I am right about 2, no argument for 1 is complete unless it makes clear why R is false. Since Quine's argument does not make clear why R is false, it fails.

Notice that the truth of 2, which is roughly what I was trying to establish in my earlier paper, does not depend on the truth of R, contrary to what Bradley seems to have assumed. Notice too that if Quine's 3 were supposed to be a premiss of his argument for Q, his whole argument

would be redundant: Q would follow trivially from 3; nor could the argument do its job, which is to persuade those who do *not* already accept the indeterminacy of translation. Nor can it be relevant to point out that Quine has an argument that will take him from Q to not–R (4), when the question is whether his present argument can get him as far as Q.

<div align="center">II</div>

Before discussing Bradley's second objection I will briefly recapitulate the main steps of my counter-argument.

A1.   Assume that the underdetermination of theory sets in with theoretical physics.

(Of course this is an artificial assumption because underdetermination is not thought of as setting in suddenly. But it is legitimate when the point is simply, in Bradley's words, 'to break Quine's argument from the fact of underdetermination to the existence of the indeterminacy'.)

A2.   Assume that translation is determinate up to theoretical physics.

(This corresponds to R above. Quine's argument fails if on this assumption there can be determinate translation of undetermined theory: A2 does not have to be *true*.)

A3.   Then the Martians might have a physical theory M such that we could make a uniquely correct partial translation of a textbook of M which revealed that M was *isomorphic* with the English physical theory A. That is, the result of applying to the textbook a word-for-word mapping of the Martian theoretical vocabulary onto the vocabulary of A is an accurate exposition of A; and similarly for all supplementary explanations of M which Martian physicists were disposed to give.

A4.   In that case the *only* differences between Martian M and English A would be ones of vocabulary.

A5.   But mere difference of vocabulary is not sufficient for difference of theory: isomorphism of the sort described is sufficient for theoretical identity.

A6.   So Quine's argument fails because, given A2, there would be no alternative to ascribing theory A to the Martians, rather than any other theory empirically equivalent to A.

Bradley's objection is that

> nothing has been done to rule out the existence of other mappings, compatible with all speech dispositions but not isomorphic in Kirk's sense, and the problem remains of deciding between the various manuals of translation generated by these further mappings (*op. cit.,* p. 19).

My reply is that the existence of other mappings, though interesting, would be irrelevant because of the way in which we are forced to interpret Quine's own main assumption of the underdetermination of physical theories. The idea is that two different physical theories can both be compatible with all possible data. As I pointed out early in my original paper (pp. 196f.), the theories must be supposed to be different in some non-trivial sense. In particular mere differences of vocabulary—differences in the mere noises or marks used—must not be allowed to count as theoretical differences, for otherwise Quine's argument would be worthless. We should have to concede for example that a theory consisting of present-day physics in English, with the words 'electron' and 'neutron' interchanged throughout, would be a *different theory* from present-day physics in English. Then Quine's main assumption would be trivially true. That alone would not matter, of course. But on this interpretation the conclusion of his argument would also be trivially true, so the argument would be superfluous. For on this interpretation, if by one system of translation we translate the foreigner as 'believing *A*', there will be a different theory, which consists of *A* with certain pairs of words interchanged throughout, which he can also consistently be translated as believing—and that is Quine's conclusion. Nor is that all. For when this trivializing interpretation is applied to the indeterminacy thesis itself, the latter has no interesting or important implications for the philosophy of language. In particular it provides no support for Quine's strictures against the objective validity of the notions of meaning, synonymy, analyticity, proposition, etc. I conclude that we must interpret Quine's main assumption in such a way that if the only differences between theory *A* and theory *B* are differences of vocabulary, then *A* is the same theory as *B*. Hence step A5 of my argument. (See also sec. IV below.)

We can now see that even if there were 'other mappings, compatible with all speech dispositions but not isomorphic in Kirk's sense', their existence would have no tendency to undermine my conclusion (A6) that the Martian theory *M* would be, given steps A1–A5, *the same theory* as our theory *A*. And if *M* is the same theory as *A*, there is no logical room to ascribe some (genuinely) different theory *B* to the Martians. So Bradley's objection fails.

III

These considerations will also help to rebut the objection raised by Peter Smith. I argued:

> (2) If English-speaking instructors can successfully teach their pupils *A* rather than *B*, then Martian ones can successfully teach ignorant translators their theory *M* rather than some other. (3) And the former must be possible, since otherwise nobody could possibly tell whether he or anyone else held theory *A* rather than theory *B*, and it would make no sense

to say that these are different theories, which in turn would make nonsense of both Quine's main assumption and his conclusion (p. 200: Smith's numbering).

Smith alleges a non-sequitur at (3). My mistake, he thinks, is to have assumed

> that the possibility of one English speaker transmitting the 'same theory' to another is a precondition of anybody telling whether he himself holds one theory rather than another: that is clearly wrong, . . . (pp. 43of.)

Now certainly I am not entitled to assume a universal principle to that effect. But my point at (3) is much more restricted. For instance it is not supposed to hold independently of the assumptions A1, A2 and A5 made in my argument. In fact it follows logically from these assumptions. For it is simply the point that on these assumptions any English-speaker who can tell whether he holds physics A rather than physics B must be able to tell that another English-speaker holds A rather than B, when the latter holds true either the sentences of A or a set of sentences isomorphic in my sense with those of A. A1 and A2 ensure that he can tell which sentences the other holds true, while A5 ensures that if the sentences held true include the sentences of A, or sentences isomorphic with those of A, then the other holds theory A. So there is no non-sequitur.

Perhaps it should be emphasized that Quine's argument cannot be saved by making a special exception of the isomorphic case. The fact that M is isomorphic with A is logically independent of the under-determination of each by the totality of observation sentences. If Quine's argument were sound, it would have to work for the case of translation between isomorphic theories just as well as for other cases.

IV

The tension between Quine's main assumption about the under-determination of theories and his indeterminacy thesis shows up in-terestingly in a recent suggestion he has made for elucidating the former. 'If two theories conform to the same totality of possible observations, in what sense are they two?' ('The Nature of Natural Knowledge', in *Mind and Language*, ed. by S. Guttenplan, Oxford, 1975, p. 80). Merely terminological differences may be disregarded (which gives us A5). Again, if the two theories are alike except that one assumes an infinite space while the other has a finite space in which bodies shrink in pro-portion to their distance from the centre,

> we want to say that the difference is rather terminological than real; and our reason is that we see how to bring the theories into agreement by translation: by reconstruing the English of one of the theories (*loc. cit.*).

On the other hand it will not do, Quine insists, to maintain that empirical equivalence—agreement with the totality of true observation sentences—renders the difference purely verbal: 'This argument simply rules out, by definition, the doctrine that physical theory is underdetermined by all possible observation' (*loc. cit.*). Quine's suggestion is:

> The best reaction at this point is to back away from terminology and sort things out on their merits. Where the significant difference comes is perhaps where we no longer see how to state rules of translation that would bring the two empirically equivalent theories together. (*Op. cit.,* pp. 80f.)

This suggestion surely has merit. But it is puzzling when juxtaposed with 'On the Reasons . . .'. For it seems to rule out, by definition, the very conclusion to which the earlier paper was directed: that where A and B are *different* physical theories both compatible with all the possible data, it might be possible for us to 'adopt A for ourselves and still remain free to translate the foreigner either as believing A or as believing B' (p. 180). For if we are free to adopt either A or B for ourselves, and can translate whatever the foreigner says in terms of either theory, presumably we can translate any sentence of either into a sentence of the other. And this must be what Quine means by 'bringing the theories into agreement by translation'. It does not matter if the correlated English sentences are not even loosely equivalent. Indeed, *unless* this were so A and B would not even appear to be different theories. Quine's suggestion is intended to deal with just such cases. What are on the face of it different theories, saying different, even incompatible things, are to be regarded as significantly different only where we do not see how to state rules of translation between them. If we do see how to do this, we thereby bring the two theories into agreement. So it seems that Quine's present suggestion would make the conclusion of 'On the Reasons . . .' trivially false. If, as his argument requires, A and B are significantly different theories, then on his present suggestion we are *not* free to adopt either theory for ourselves and translate the foreigner either as believing A or as believing B. If they are different theories, we cannot translate the foreigner as holding more than one of them—so there is no room for indeterminacy.[1]

---

[1] In correspondence Professor Quine has agreed that the point from 'The Nature of Natural Knowledge', which he has developed more fully in 'On Empirically Equivalent Systems of the World', *Erkenntnis,* 9 (1975), 313–328, does demolish the argument from A and B that he had proposed in 'On the Reasons for Indeterminacy of Translation'.

University of Nottingham   © ROBERT KIRK 1977
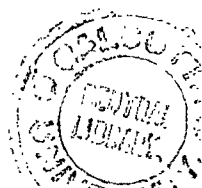
PRIMA FACIE AND ACTUAL DUTY

By Arthur M. Wheeler

THE criticisms of ethical intuitionism by Strawson and Nowell-Smith, among others, have gained considerable acceptance. Some philosophers have even found ethical intuitionism incredible. The issue is far from dead; discussion and criticism continue in recent books by W. D. Hudson, G. J. Warnock, Mary Warnock, and others. On the other hand, an occasional article in general defence is written. My argument will be to the more restricted point that some intuitionist views are not so naïve as some critics take them to be.

Recently, in *Moral Philosophy* (Macmillan, 1967), Richard Garner and Bernard Rosen have produced an objection to W. D. Ross which at first blush seems convincing, but which can be answered. In their main account of Ross (pp. 104–109) they give a counter-example to Ross's *prima facie* duty to make better the condition of others in respect of virtue, intelligence and pleasure. 'We do not have a prima facie duty to tell our neighbour intimate details concerning our love life, even though that would increase his knowledge and perhaps even give him some pleasure. In fact, we seem to have a duty not to increase his knowledge or pleasure in this way. That is, Source 4 is activated positively, none of the other sources is activated at all, and yet it is false that we have a prima facie duty to do something; in fact we seem to have a duty not to perform that action.'

I believe that in the last sentence the argument moves unfairly from *prima facie* duty to actual duty. I think Ross would agree, generally, that we have an *actual* duty to *not* increase the neighbour's knowledge or pleasure in this way. He could still hold there is a *prima facie* duty to inform the neighbour, but it would hardly ever be one's *actual* duty. Its being an actual duty would be ruled out by the special relation between the person and his lover.

In his rejection of ideal utilitarianism, in chapter 2 of *The Right and the Good,* one of Ross's complaints is that the theory simplifies unduly our relations to our fellows. He lists a number of relations, among which is the relation of wife to husband, and says that 'each of these relations is the foundation of a *prima facie* duty, which is more or less incumbent on me according to the circumstances of the case' (p. 19). Since Garner and Rosen urge that we suppose our partners in our love life are all casual, I shall expand Ross's account to include the lover–loved, or even lover–acquaintance, relation. I assume Ross would allow this, since he does not claim to give an exhaustive list of *prima facie* duties, or, I suppose, of relations to others.

When Garner and Rosen say that 'none of the other sources is

activated at all', they are dealing only with the *prima facie* duties Ross specifically lists. But Ross has taken care of the case they give by saying we have special duties to others, depending on our relation to them.

Their complaint should be that the special relationships Ross mentions need to be repeated in his statement of the particular *prima facie* duties; otherwise we shall forget them and will mistakenly consider only the particular *prima facie* duties he lists.

In summary, on Ross's view we could argue that there is a *prima facie* duty to increase the knowledge and pleasure of the neighbour by telling one's love life. But this *prima facie* duty does not, except, perhaps, in very unusual circumstances, become an *actual* duty. This is so because its becoming an actual duty is ruled out by the special relation of lover–acquaintance.

We can also conceive of counter-examples to Garner and Rosen's account. If we press Ross's *prima facie* duty of beneficence, we can think of cases where Ross might even allow it is the *actual* duty for a *certain* person so to inform his neighbour; e.g. the case of a psychologist whose neighbour is his patient. The patient's marital life is disastrous, and the situation is one of group therapy, in which revelation of a personal nature is needed to increase the neighbour's trust in the psychologist. Or, perhaps the psychologist feels that a bit of such information about his own personal life will *greatly* improve the knowledge (let's omit the issue of pleasure) of the neighbour in a way which is very important for the neighbour's marital life, and indeed perhaps, in the judgement of the psychologist, even his sanity and continued existence. On Ross's principles, if, for example, the good results will *far* outweigh the keeping of a promise (which I take to be involved in the husband–wife, or in this case, the lover–acquaintance, relation), then the greater duty, or his *actual* duty, would be to break the promise. I should add that my suggestion is given in the context of Ross's view that any judgment of *actual* duty is fallible.

I certainly agree that my counter-example is an extreme case; I agree that in the vast majority of cases a person should *not* divulge these things to his neighbour. But would Garner and Rosen want to claim *no* one would *ever* have the duty so to inform his neighbour, for example, even to save a life? If they would, they would be as rigid as Kant was, in his famous rejection of telling a lie to save a life. (I am aware that Dietrichson, for example, thinks this an atypical passage in Kant; see Paul Dietrichson, 'Kant's Criteria of Universalizability', in *Kant, Foundations of the Metaphysics of Morals,* ed. R. P. Wolff, Bobbs-Merrill, 1969, p. 206; cf. *Kant-Studien,* 1964, pp. 143–70.)

Of course there might be countries where it is moral and even flattering to the partner to tell others one's love life. The issues are not settled, I think, but such cultural differences *are* usually cited in the attack on

intuitionism. Perhaps, although Ross attempts it, intuitionists cannot adequately take care of such considerations, since it might be argued that if people of different countries or communities intuitively 'see' things differently, this mere fact, if it is a fact, undermines intuitionism as a method of gaining 'truth'. But such argument need not be met in order to meet the particular complaint of Garner and Rosen, for it is clear that their criticism is within the *assumed* moral outlook of Anglo-American or Western European society. That context surely *assumes*, generally, that one ought, in the moral sense, to be silent in matters such as one's love life; at least, on their argument, that would also be assumed by Garner and Rosen. And on that basis, my argument meets their criticism.

Kent State University   © ARTHUR M. WHEELER 1977

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Christopher Kirwan, Exeter College, Oxford OX1 3DP. Contributors are asked to note the following requirements.

Articles should normally not exceed 3,000 words in length. Occasionally, however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope and international reply coupons of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

REPORT ON *ANALYSIS* "PROBLEM" NO. 15

By BERNARD WILLIAMS

There exist just 100 tablets in a certain script, as yet undeciphered. Scholar A considers all the tablets, and works out a decipherment which makes sense of them. Independently, scholar B selects 50 of the tablets, and from them works out a decipherment which he then tries out, successfully, on the other 50. Is one of these scholars behaving more rationally than the other? If their decipherments differ, does the difference in procedure give any reason to accept one rather than the other?

THE competition attracted only eight entries. Various entrants argued, equally confidently, that A's strategy was the more rational; that B's was; that it depended on the circumstances; and that it could make no difference, the last itself being urged for more than one reason. Most competitors took the point that it was meant to be a puzzle about hypothesis-testing in general, and not specifically about decipherment.

Popper was mentioned (rather hopefully) by more than one entrant, but no-one raised a question suggested by the puzzle for a Popperian philosophy of science, namely how an event of hypothesis-testing is to be identified, and, connectedly, how a notion of corroboration which is partly an historical notion is to be applied.

Mr D. A. Osborn, however, did make an important point relevant to this issue: that if one actually considers how A must proceed, it is quite unclear how different in principle his procedure can be from B's. Mr Osborn uniquely made this point, and developed some of its implications, and for this reason I recommend that he be awarded the prize. At the time of submitting his entry, Mr Osborn was a first-year student in philosophy and mathematics.

ON DECIPHERING THE TABLETS

By D. A. OSBORN

THE problem cannot be tackled without making a few assumptions. One must first assume that there exists at least one coherent intelligible decipherment and that, if there is, then it can be understood with our language and conceptual resources. One must also assume that there are distinguishable and recognizably systematic characters on the tablets in some kind of order. Another fundamental assumption

must be made as to whether there is only one, or more than one, distinct consistent decipherment.

Firstly, if there is only one decipherment for which all the tablets are consistent then the last question is inapplicable. In considering the 100 tablets Scholar A is not going to be able to consider them all simultaneously (through practical and physical limitations) and so he will have to consider first one—and make it cohere with a second. He would then have to make these compatible with another two and so on. He would continue checking more and more against each other—keeping on making the necessary alterations. This procedure *may* involve considering fifty and checking with a further fifty. However he wouldn't exclusively deal with an arbitrarily chosen fifty tablets because of the possibility of other characters appearing in the fifty not initially considered, which have not been encountered in the first fifty. Also, by the very nature of the method of decipherment he would still have to be prepared to make the necessary alterations in the other fifty. If there were characters not previously met in the second fifty then by dealing exclusively with one lot of fifty his method would be useless (that is, assuming that not every character appears on every tablet). So the rational scholar while not, for practical reasons, considering *all* 100 tablets as such, will equally not study fifty exclusively. But apparently Scholar B succeeds, which can only be due to mere chance and not rational calculation. There is no reason why a randomly selected set of fifty tablets considered on their own should cohere with the rest.

Secondly, if there is more than one different decipherment which separately makes sense one's immediate reaction would be to accept Scholar B's decipherment. But as I have pointed out Scholar A has to use this kind of technique in considering the 100. If it comes to the stage when he has collated 50 tablets and ensured their consistency he will compare then with the others. If the whole lot doesn't work then he can modify the decipherment until it does—presumably along fairly similar lines. If it does then it approximates to the case of Scholar B.

In both cases I think it is clear that the more rational way is that taken by Scholar A of which Scholar B's approach is a special case. Neither can really be considered as more acceptable than the other from within the system, without considering some extra external information or evidence. This is because both Scholars are essentially doing the same thing although because of the possibility of different decipherments they might reach different conclusions.

University of Bristol   © D. A. OSBORN 1977

DISTURBANCES

By Toomas Karmo

A STREAM of water, e.g. the stream running down a rain-soaked windscreen, must be distinct from the water which it at any given moment happens to contain, for one and the same stream may at two distinct moments contain two distinct consignments of water. But if water and stream are distinct, then in what does that relation consist which we describe by saying that the one constitutes the other?

The correct answer to this is perhaps that a stream is a species of disturbance, where a disturbance is definable as an object or entity found in some other object—not in the sense in which a letter may be found in an envelope, or a biscuit in a tin, but in the sense in which a knot may be in a rope, a wrinkle in a carpet, a hole in a perennial border, or a bulge in a cylinder. One way of telling whether an object X is "in" an object Y in the sense peculiar to disturbances is to enquire whether X can migrate through Y. A knot is a disturbance because it may slip along the rope in which it is tied, and a hole a disturbance because it can be pictured as moving around the flowerbed in which it was dug (as a vortex may move in a pool). A bulge in a cylinder, again, is a disturbance, because it can be pictured migrating in the way a bulge migrates down the body of a boa constrictor when it swallows a goat. That which a disturbance is in is its medium; a stream is a disturbance in that total consignment of water which is now, has at any time in the past been, or will at any time in the future be found in it. The process which is the flowing of a stream may equally well be described as a stream's migrating through a quantity of water, even though in an absolute sense the stream stays stationary and the water does the migrating. (The stream migrates in the sense in which a kink migrates along a hose pulled through a stationary pair of constricting rollers.)

Streams are not alone in being constituted of distinct consignments of stuff at distinct instants in their careers. A living creature can be conceived of as a disturbance migrating through a consignment of organic chemicals, since the consignment of organic stuffs constituting a living thing one month is distinct from the consignment constituting it the next. Those, again, who suggest that one and the same pair of stockings could be first silk and then, at the end of a long process of patching, worsted in effect suggest that stockings be viewed as disturbances in consignments of fibre.

It is possible that this view of living creatures as disturbances was held by Aristotle. At any rate Aristotle mentions an object which we would call a disturbance while discussing substantiality in the *Metaphysics*: '... what is still weather?—A stillness in a large expanse of air.

The matter is the air, while the actuality and substance is the stillness' (1043a22–24). He uses this same preposition 'in' in describing the relation of the form of a man to his flesh and bones ('*to toionde eidos en taisde tais sarxi kai ostois*', *Metaphysics* 1034a6), and explicitly compares living tissues with flowing water at *de Gen. et Corr.* 321b24–25. Finally, if Aristotle does not discuss the relation of streams as such to the water in them, he does at least suggest at *Topics* 127a3–8 that a wind is better defined as movement in air than as air in movement; and the parallel between wind and air on the one hand and stream and water on the other is fairly obvious.

St. John's College, Oxford   © Toomas Karmo 1977

A NOTE ON MR. KARMO'S DISTURBANCES

By T. J. M. Bench-Capon

IN his essay on disturbances, Toomas Karmo asks us to believe three things:

(1) That X may be in Y in two quite different ways, exemplified by 'a biscuit is in the tin' and 'a dent is in the tin',

(2) that a way of telling if X is in Y in this second way is to see whether it is possible for X to migrate through Y,

(3) that streams are in consignments of water and living organisms in consignments of chemicals in this second way.

I want to argue that if (2) is true, then (3) is false, and the reason why this is so should lead us to reject (1) also.

If X is to be able to migrate through Y, then it must make sense to ask where X is in Y, because the possibility of X's migrating through Y depends on the possibility of there being different answers to this question at different times. But it makes no sense to ask where the stream is in the water, or where the organism is in the chemicals. So, they cannot migrate in the required way and so fail the test advocated in (2).

Moreover, the fact that it does make sense to ask where the X is in the Y in both the paradigms offered in (1) ought to lead us to suspect that these ways of being in are not as disparate as Mr Karmo would have us believe. For both these ways of being in, that X is in Y entails that X is somewhere in Y. This entailment does not hold for streams and water or organisms and chemicals, and so if we want to hold on to the claim that

streams are in water, we would want a way of being in where this entailment does not hold. There is one; the way in which a pattern may be in a carpet. But patterns are not disturbances in carpets (or at least not in the way in which dents are disturbances in tins) and this would cast doubt on the view that streams are disturbances in consignments of water, and that organisms are disturbances in consignments of chemicals.

St. John's College, Oxford   © T. J. M. BENCH-CAPON 1977

A CAUTION ON PROPOSITIONAL IDENTITY

By JAMES B. FREEMAN

IN their article, 'Prior on Propositional Identity' (ANALYSIS, 36. 4), Philip Hugly and Charles Sayward claim that the following 'is arguably a logical truth':

(1) $\forall p \forall q \,((\delta p = \delta q) \supset (p = q))$.

The authors realize that counterexamples may be constructed where δ is taken as a tautologous function (e.g., where δp is '$p \vee \sim p$'), if the assumption be granted that all tautologies are identical. However, they feel that the desirability of (1) as a logical truth gives strong justification for rejecting the thesis that all tautologous propositions are identical.

I want to make two points: (a) there are intuitive reasons for connecting the notions of propositional identity and propositional entailment or implication[1]; (b) anyone who accepts this connection will be committed, if he also accepts intuitively plausible entailment or implication principles and standard logical principles, and furthermore accepts (1), to holding the patent falsehood that all propositions are identical.

Connecting propositional entailment and propositional identity means accepting at least one of the following principles: that two propositions are identical when they imply each other, or when they entail each other, or when they provably imply, or entail, each other. For instance, in *Meaning and Necessity* (Chicago and London: The University of Chicago Press, 1947), Carnap claims that frequently a proposition is understood as what may be expressed by a declarative sentence. Two

---

[1] However, Professor Sayward has informed me that neither he nor Professor Hugly identify propositional identity and propositional equivalence, and believe (1) should be given up if such an identification is made.

propositions are identical just in case the sentences expressing them are
L-equivalent. This connects propositional identity and L-equivalence,
and motivates the claim that some such connection is correct. Also,
frequently in syntactically investigating a system of modal logic, one
shows that if $\vdash A \equiv B$, $\vdash E[A] \equiv E[B]$, where E is some context in
which A or B may occur. But this is to argue, at least for the system in
question, that A and B are indiscernible. Within the language, there is no
context which distinguishes A and B. Accepting identity of indiscern-
ibles, such a proof attempts to establish that for the system in question
provable equivalence amounts to propositional identity.

Prior would vigorously deny any identification of propositional
identity with some type of propositional equivalence. The intuitive
motivation here comes from taking propositional attitude contexts as
genuine modalities of propositions, e.g., 'Jones believes that — — —'.
Given that identical propositions are indiscernible, clearly two pro-
positions might be equivalent, say 'p' and '$\sim\sim p$', but yet Jones might
believe that p and fail to believe that $\sim\sim p$. On the above assumptions,
this would entail that 'p' and '$\sim\sim p$' are distinct, although logically
equivalent, propositions. Propositional attitude contexts do pose a
problem for anyone attempting to identify propositional identity with
some form of equivalence. However, I do not believe that these pro-
blems are insurmountable, although I cannot discuss this point here.
What I want to urge is that if one wants to make this identification and
also hold principle (1), incautious acceptance of intuitively plausible
entailment principles will lead to the disastrous consequence that all
propositions are identical. I turn now to establishing this point.

If one identifies propositional identity with provable material equi-
valence, or provable relevant equivalence in an Anderson-Belnap style
relevant logic, then where '$\leftrightarrow$' expresses either one of these types of
equivalence,

(2) $p \leftrightarrow (p \vee (p \wedge q))$.[1]

Hence

(3) $p = (p \vee (p \wedge q))$

and likewise

(4) $p = (p \vee (p \wedge r))$.

Since '$\vdash p = p$' is indisputably a principle of propositional identity,
by (1) (understood with '$\rightarrow$' replacing '$\supset$') we have that $\vdash q = r$ for
arbitrary propositions.

---

[1] I want to thank both Professors J. Michael Dunn and Hector-Neri Castañeda for
suggesting this example to me, and Professor Dunn for suggesting this use of '$\leftrightarrow$'.

Notice that 'p' is not a tautology. Hence the mere avoidance of the identity of all tautologies (which is effected in relevance logic) does not free a system incorporating (1) from identifying all propositions.

In fact, we may go further. Any theory of implication accepting the principles

(5) $\vdash p \to (p \vee q)$
(6) $\vdash ((p \to r) \wedge (q \to r)) \to ((p \vee q) \to r)$

and set within a logical system incorporating ordinary principles of quantification for propositional variables, has as theorems

(7) $\vdash (\exists p\, p \vee q) \leftrightarrow \exists p\, p$

and

(8) $\vdash (\exists p\, p \vee r) \leftrightarrow \exists p\, p.$[1]

But this again gives us that $\vdash q = r$.

However, not all theories of implication accept (5), for example, Parry's analytic implication. But analytic implication does accept

(9) $\vdash (p \wedge q) \to p.$

Yet given (9) together with

(10) $\vdash ((p \to q) \wedge (p \to r)) \to (p \to (q \wedge r)),$

we may invert the above argument to shown that $\vdash (\forall p\, p \wedge q) = \forall p\, p$ and $\vdash (\forall p\, p \wedge r) = \forall p\, p$, yielding again that $\vdash q = r$. Hence any theory which accepts (1) and identifies propositional identity with equivalence in some sense, must reject either (9) or (10) as characterizing the corresponding implication, or must alter the laws of quantification when applied to propositions.

In the light of the above problems, one doubts whether (1) may be retained when identifying propositional identity and equivalence. In addition, although Hugly and Sayward can intuitively motivate (1) with examples, what (1) says in effect is that all functions of propositions are one–one. But this is clearly not intuitive.[2]

University of Victoria   © JAMES B. FREEMAN 1977

KRIPKE ON THE A PRIORI AND THE NECESSARY

By ALBERT CASULLO

PHILOSOPHERS have traditionally believed that there is a close connection between the categories of a priori propositions and necessary propositions. One widely held thesis about the nature of this connection is that all a priori knowledge is of necessary propositions and that all necessary propositions are knowable a priori.[1] Saul Kripke has recently argued that this traditional account is mistaken. In 'Identity and Necessity'[2] he argues that there are necessary a posteriori propositions, while in 'Naming and Necessity'[3] he argues, in addition to this, that there are contingent a priori propositions. The primary concern of this paper is to examine Kripke's arguments in order to determine whether he has succeeded in calling the traditional account into question.

I

Kripke's claim that there are necessary a posteriori propositions arises in the context of a discussion of essential properties. He begins with the following consideration:

> Supposing this lectern is in fact made of wood, could this very lectern have been made from the very beginning from ice, say frozen from water in the Thames? One has a considerable feeling that it could *not*, though in fact one certainly could have made a lectern of water from the Thames, frozen it into ice by some process, and put it right there in place of this thing. If one had done so, one would have made, of course, a *different* object.[4]

Therefore, in any counterfactual situation in which this lectern existed, it would not have been made from water from the Thames frozen into ice. Kripke goes on to argue that if the essentialist view is correct, then there are necessary propositions knowable only a posteriori. He summarizes his argument in the following manner:

> In other words, if *P* is the statement that the lectern is not made of ice, one knows by a priori philosophical analysis, some conditional of the

---

[1] For example, Kant states in the *Critique of Pure Reason*, trans. Norman Kemp Smith (New York: St. Martin's Press, 1965), p. 11, that 'Any knowledge that professes to hold *a priori* lays claim to be regarded as absolutely necessary'. Leibniz claims in *The Monadology* that 'There are also two kinds of *truths*, those of *reasoning* and those of *fact*. Truths of reasoning are necessary and their opposite is impossible, and those of *fact* are contingent and their opposite is possible. When a truth is necessary its reason can be found by analysis, resolving it into more simple ideas and truths until we reach those which are primitive.' See *Leibniz: Selections*, ed. P. P. Wiener (New York: Charles Scribner's Sons, 1951), p. 539.

[2] Saul A. Kripke, 'Identity and Necessity', in *Identity and Individuation*, ed. M. K. Munitz (New York: New York University Press, 1971).

[3] Saul A. Kripke, 'Naming and Necessity', in *Semantics of Natural Language*, eds. D. Davidson and G. Harman (Dordrecht: D. Reidel Publishing Company, 1972).

[4] *Identity and Individuation*, p. 152.

form "if P, then necessarily P". If the table is not made of ice, it is necessarily not made of ice. On the other hand, then, we know by empirical investigation that P, the antecedent of the conditional, is true—that this table is not made of ice. We can conclude by *modus ponens*:

$$P \supset \Box P$$
$$P$$
$$\overline{}$$
$$\Box P$$

The conclusion—'$\Box P$'—is that it is necessary that the table not be made of ice, and this conclusion is known a posteriori, since one of the premisses on which it is based is a posteriori.[1]

Therefore, since presumably what a certain lectern is made of can be known *only* a posteriori, the essentialist view can be accommodated only if one rejects the thesis that all necessary propositions are knowable a priori.

The claim that if there are essential properties then there are necessary propositions which are knowable only a posteriori is ambiguous, and two different interpretations of it must be distinguished. In order to see this ambiguity, one must distinguish between knowledge of the truth value of a proposition and knowledge of its general modal status. One has *knowledge of the truth value* of a proposition when one knows whether it is true or false. One has *knowledge of the general modal status* of a proposition when one knows whether it is a necessary proposition or a contingent one. Letting 'Fa' stand for the proposition 'a has the property F' where F is an essential property of a, one can now see that the claim that there is only a posteriori knowledge of necessary propositions such as 'Fa' can be interpreted in either of the following two ways: (1) 'Fa' is knowable only a posteriori, and 'Fa' is a necessary proposition, or (2) that 'Fa' is a necessary proposition is knowable only a posteriori. If there are essential properties, then it would follow that one could have only a posteriori knowledge of the *truth value* of necessary propositions such as 'Fa'. But, even if there are essential properties, it would not follow that there can be, let alone can only be, a posteriori knowledge of the *general modal status* (or necessity) of propositions such as 'Fa'.

The claim that even if there are essential properties, it would not follow that there can be only a posteriori knowledge of the *general modal status* of some necessary propositions, might seem to be in conflict with Kripke's conclusion. For he maintains that if there are such properties, then '$\Box P$' is knowable only a posteriori. Therefore, we seem to have a case of a posteriori knowledge of the necessity (or general modal status) of a proposition. This is not so, however. One must distinguish between

---

[1] Ibid., p. 153.

the general modal status and the specific modal status of a proposition. By the *general modal status* of a proposition I mean its being necessary or its being contingent, regardless of whether it is necessarily true or necessarily false, or contingently true or contingently false. By the *specific modal status* of a proposition I mean its being necessarily true, necessarily false, contingently true, or contingently false. One must also recognize that knowledge of the specific modal status of a proposition consists of knowledge of its general modal status together with knowledge of its truth value. Hence, in cases where one's knowledge of *both* the general modal status of a proposition and its truth value is a priori, knowledge of its specific modal status would also be a priori. But in cases where one's knowledge of the truth value of a proposition is a posteriori, knowledge of its specific modal status would be a posteriori, even if knowledge of its general modal status is a priori.

Kripke's claim that '☐P' is knowable only a posteriori is a claim about knowledge of the specific modal status of a proposition and is based on the fact that where 'P' is a proposition about a physical object possessing a property, its truth value is knowable only a posteriori. But Kripke clearly does not deny that knowledge of the general modal status of propositions about essential properties is a priori. (He says that 'if P is the statement that the lectern is not made of ice, *one knows by a priori philosophical analysis,* some conditional of the form "if P, then necessarily P" '.)[1] Therefore, the existence of essential properties would entail that there are necessary propositions whose truth value and specific modal status are knowable only a posteriori, but it would not entail that there are necessary propositions whose general modal status is knowable only a posteriori.

The question whether Kripke's claims about knowledge of propositions such as '*Fa*' conflict with the traditional account of the relationship between a priori and necessary propositions is a difficult one to answer, since its proponents did not distinguish between the truth value, specific modal status, and general modal status of a proposition. We can conclude that if there are essential properties like those suggested by Kripke, then it would be incorrect to maintain that the truth value of all necessary propositions can be known a priori. From this it follows that it would also be incorrect to maintain that the *specific* modal status of all necessary propositions can be known a priori. But the existence of such properties would not call into question the claim that the *general* modal status of all necessary propositions can be known a priori.

II

In 'Naming and Necessity' Kripke attempts to strengthen his claim that the traditional account of the relationship between a priori and

---

[1] Ibid. The emphasis is mine.

necessary propositions is mistaken by providing an example of a proposition which is contingent but knowable a priori. His discussion begins with a consideration of Wittgenstein's comments about the standard metre. Wittgenstein claimed, 'There is *one* thing of which one can say neither that it is one metre long, nor that it is not one metre long, and that is the standard metre in Paris.'[1] Kripke disagrees: 'If the stick is a stick, for example, 39.37 inches long (I assume we have some different standard for inches), why isn't it one metre long?'[2] He then goes on to raise the question whether the proposition that the standard metre is one metre long at time t_0 is a necessary truth. He argues that the proposition is not necessary even if one grants that by definition the standard metre is one metre long at time t_0 because

> the 'definition', properly interpreted, does *not* say that the phrase 'one metre' is to be *synonymous* (even when talking about counterfactual situations) with the phrase 'the length of S at t_0' but rather we have *determined the reference* of the phrase 'one metre' by stipulating that 'one metre' is to be a *rigid* designator of the length which is in fact the length of S at t_0. So this does *not* make it a necessary truth that S is one metre long at t_0.[3]

Kripke goes on to claim that for a person who fixes the metric system by reference to stick S at t_0, the proposition 'Stick S is one metre long at t_0' is known a priori:

> For if he used stick S to fix the reference of the term 'one metre', then as a result of this kind of 'definition' (which is not abbreviative or synonymous definition), he knows automatically, without further investigation, that S is one metre long.[4]

Therefore, the proposition that stick S is one metre long at t_0 is both contingent and knowable a priori.

In order to evaluate this argument we must distinguish the following two sentences: (1) S is one metre long at t_0; (2) The length of S at t_0 is one metre. A further distinction must also be made between two possible interpretations of the second sentence. We may follow Donnellan by pointing out that the definite description 'the length of S at t_0' can be used either attributively or referentially.[5] (I shall not attempt to defend

---

[1] Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1968). p. 25.
[2] *Semantics of Natural Language*, p. 274.
[3] Ibid., p. 275.
[4] Ibid.
[5] See Keith S. Donnellan, 'Reference and Definite Descriptions', *The Philosophical Review*, LXXV (1966): 281–304. On page 25 he states,
> A person who uses a definite description attributively in an assertion states something about whoever or whatever is the so-and-so. A speaker who uses a definite description referentially in an assertion, on the other hand, uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or thing.

the distinction here.) When a speaker uses the sentence, 'The length of S at t_0 is one metre,' to introduce the term 'one metre', he might be making either of the two following claims: (a) he wishes to introduce 'one metre' as the name of the length of S at t_0 *whatever* that length might be; (b) there is a particular length which he has in mind and which he can identify independently of the truth of the proposition that it is the length of S at t_0, and it is *this* length which he wishes to call 'one metre'. Depending on how the definite description is used to introduce the term 'one metre', what is asserted by (2) and, consequently, also what is asserted by (1) will change.

If one uses the definite description attributively in introducing 'one metre' by means of sentence (2), then one is using 'one metre' as the name of the length of S at t_0 whatever it may be. The term is not being introduced as the name of a particular length which the speaker has singled out but as the name of whatever length happens to satisfy the definite description. This method of introducing the term results in what Kripke calls an 'abbreviative definition', for the speaker is using the term 'one metre' as an abbreviation for the phrase 'the length of S at t_0'. As a result of this definition, the proposition expressed by the sentence 'The length of S at t_0 is one metre' is a necessary one, true solely in virtue of the terms used in expressing it. Since it is true solely in virtue of the meanings of its terms, it is also knowable a priori. If the term 'one metre' is introduced in this manner, the proposition expressed by the sentence 'S is one metre long' is also a necessary one. Since 'one metre' is an abbreviation for 'the length of S at t_0', the proposition expressed by the sentence 'S is one metre long' is identical to the one expressed by the sentence 'S has at t_0 whatever length it does have at t_0' which is trivially true. Hence, if the term 'one metre' is introduced by means of a sentence which uses the definite description attributively, both propositions—that expressed by the sentence 'The length of S at t_0 is one metre' and that expressed by the sentence 'S is one metre long at t_0' —are necessary and a priori.

The situation is not the same, however, if one uses the definite description referentially in introducing the term 'one metre' by means of sentence (2), for the speaker is not introducing 'one metre' as the name of whatever length happens to satisfy the definite description 'the length of S at t_0'. Instead, he is introducing it as the name of a *particular* length to which he tries to call attention by using the definite description. He uses this particular definite description because he believes that S in fact has the length he wishes to name. But if it should happen that due to some peculiar environmental conditions S does not have the length he thought it had, then the speaker would have introduced 'one metre' as the name of the length which he thought S had, rather than the one which it in fact had. Therefore, the term 'one metre' is not being used as a

synonym for 'the length of S at t_0' but as the name of a particular length, whether or not it is in fact the length S at t_0.

Since this point might not be clear in the case of lengths, let us consider the case of colours. Suppose someone were to introduce the term 'red' using the definite description 'the colour of S at t_0' referentially. Also, suppose that he is using the definite description to refer to the colour red; but, because of some peculiar lighting conditions unknown to the speaker and everyone else in the immediate vicinity of S, although S appears red, it is in fact white. Since the speaker was using the definite description to draw attention to a particular colour, and it was that particular colour he wished to name 'red', he would have introduced 'red' as the name of the colour red despite the fact that the colour satisfying the definite description 'the colour of S at t_0' was white. In such a case, a necessary proposition does result in virtue of the definition of 'red'. This necessary proposition, however, is not satisfactorily expressed by the sentence 'The colour of S at t_0 is red'. It is more accurately captured by the sentence, '*This* colour is red', where 'this colour' refers to the colour the speaker singled out using the definite description. This proposition is also knowable a priori, since it can be known solely on the basis of the definition of 'red'.

Returning to our original example, introducing 'one metre' as the name of a particular length to which one calls attention with the definite description 'the length of S at t_0' also yields a necessary proposition, which can be best expressed by the sentence '*This* length is one metre', where 'this length' refers to the length to which the speaker was calling attention. This proposition is also knowable a priori. But this is not true of the proposition expressed by the sentence 'S is one metre long'. Since 'one metre' has been introduced as the name of a particular length, the proposition expressed by the sentence 'S is one metre long' is no longer identical to the one expressed by the sentence 'S has whatever length it does have'. Instead, what it asserts is more accurately expressed by the compound sentence 'S has this length (rather than another), and this length is one metre'. As was stated above, the second conjunct is both necessary and knowable a priori. But this is not true of the first conjunct. For, as Kripke correctly points out, it is a contingent fact about S that it has any particular length; had the environmental conditions been different at t_0, S would have had a different length at t_0. We must notice, however, that this conjunct is also knowable only a posteriori. For although one knows a priori that the length one singled out with the definite description 'the length of S at t_0' is one metre, one does not know a priori that S *in fact* has that length. One can know this only on the basis of a posteriori considerations, such as the manner in which the object appears and the conditions under which it appears in that way. Therefore, the sentence 'S is one metre long at t_0' expresses a

contingent and a posteriori proposition when 'one metre' is introduced by means of a sentence which uses the definite description 'the length of S at t_0' referentially.

Let us consider again our example of the speaker who introduces the term 'red' using the definite description 'the colour of S at t_0' referentially. Although he knows a priori that *this* colour is red, he does not know a priori that S is red, for he does not know a priori that S has the colour he named 'red'. If he were to infer that S was red on the basis of the manner in which he introduced the term 'red', not only would he be unjustified, but he would also be mistaken in this case, since, by hypothesis, S is in fact white at t_0. He would be justified in believing that S is red at t_0 only if he knew that S appears red and that only red objects appear red under the conditions in which S appears red. But both of these facts can be known only a posteriori. Therefore, his knowledge that S is red is a posteriori. It is based on his knowledge that S has this particular colour (rather than another), which is a posteriori, and his knowledge that this colour is red, which is a priori.

It might be argued that there is a third way of introducing the term 'one metre' which has not been considered.[1] When a speaker uses the definite description 'the length of S at t_0' attributively in introducing 'one metre', there are two possibilities: (1) the speaker might be using the definite description to give the meaning of the term, in which case 'one metre' is an abbreviation for 'the length of S at t_0'; or (2) the speaker might be using the definite description to fix the reference of the term, in which case 'one metre' is the name (or rigid designator) of whatever length S happens to have at t_0. Although we have considered the first case, we have neglected the second. The primary reason for neglecting this case is that it does not constitute a genuine possibility. Since, by hypothesis, the description is not used referentially, how can one generate a genuine name from it? How can one, relying solely on the description, provide the term with reference? The appeal to the vague and unexplained notion of 'fixing reference' does not by itself provide answers to these questions. If the term is to be a genuine name, rather than merely an abbreviation of the description, there must be criteria for its use which are not simply the criteria for the use of the description. It must be possible, at least in principle, for someone to determine whether the term is used correctly on future occasions without relying on the description. (This, of course, would be possible if the description had been used referentially.) If this is not possible, then it is no longer clear in what sense the term is *not* a mere abbreviation of the description and the distinction between the term's being a name and its being such an abbreviation appears to be of little consequence. Therefore, at the very least, Kripke owes us much further explanation.

---

[1] This point is due to the editor.

The failure to provide a convincing example of a contingent a priori proposition removes the basis of Kripke's second argument against the traditional account of the relationship between the a priori and the necessary. He has not given us any reason to suppose that the traditional philosophers were mistaken in claiming that all a priori knowledge is of necessary propositions.[1]

---

[1] I am indebted to Professor Panayot Butchvarov for a number of illuminating discussions on several aspects of this paper.

University of Colorado   © ALBERT CASULLO 1977

THE ATTRIBUTIVE USE OF PROPER NAMES

By A. P. MARTINICH

IT is widely held that Keith Donnellan has successfully argued that definite descriptions in the subject position of a sentence have, among their possible uses, an attributive use, which is distinct from the philosophically familiar referential use. Here is his complete characterization of it:

> A speaker who uses a definite description attributively in an assertion states something about whoever or whatever is the so-and-so. A speaker who uses a definite description referentially in an assertion, on the other hand, uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or thing. In the first case the definite description might be said to occur essentially, for the speaker wishes to assert something about whatever or whoever fits that description; but in the referential use the definite description is merely one tool for doing a certain job—calling attention to a person or thing—and in general any other device for doing the same job, another description or a name, would do as well. In the attributive use, the attribute of being the so-and-so is all important, while it is not in the referential use ('Reference and Definite Descriptions', *Philosophical Review* 75 (1966), p. 235).

In this paper I want to argue that if we suppose that there is an attributive use and that Donnellan has adequately characterized it, then there is also an attributive use of ordinary proper names. In itself,

this would be an important result since Donnellan does not think that proper names have an attributive use; indeed, a major motivation for the distinction between attributive and referential uses is just to distinguish proper names from descriptions (see Donnellan's 'Proper Names and Identifying Descriptions', in *Semantics of Natural Languages*, ed. Donald Davidson and Gilbert Harman). However, if proper names do not have an attributive use, my argument shows that the original supposition is false, that is, either there is no attributive use or Donnellan has failed to characterize it satisfactorily.

My strategy is simple. I shall present an instance of a use of a proper name that satisfies Donnellan's characterization of the attributive use (if anything does); this instance will be such that the name will be used to state something about whoever or whatever is so named; the attribute of being so named will be all important; and the name will occur essentially.

In order to know that my instance does satisfy Donnellan's characterization, it is important to know what he means by an essential occurrence of an expression. Since Donnellan unfortunately does not explain what he means by this notion, it is necessary for us to try to provide at least some explanation of it. *Prima facie*, it might seem that an expression occurs essentially if and only if no other expression can replace it. This attempt however is wide of the mark. Since any expression can be replaced by any other, no expression would occur essentially. What we need to do is to specify as best we can a law that will license certain substitutions on essentially occurring expressions. The passage quoted from Donnellan provides a suggestion to this end. After saying that definite descriptions used attributively 'occur essentially', he contrasts this with the referential use, in which 'the definite description is merely one tool for doing a certain job, and in general any other device for doing the same job, another description or a name, would do as well'. The suggestion is that for descriptions occurring essentially no other device can do the same job; that is why the original description is essential. Generalizing this suggestion to all expressions, we might say, 'An expression occurs essentially just in case no other expression can do the same job'.

But this condition also seems too restrictive because it in fact fails to license many substitutions that Donnellan would have to allow in order to preserve his claim that certain descriptions occur essentially. For example, he gives a scenario in which 'Smith's murderer' is supposedly used attributively ('Reference and Definite Descriptions', pp. 285–286). Since 'the felonious killer of Smith' does the same job as 'Smith's murderer', Donnellan would have to license the substitution of such synonymous expressions. Otherwise he would be forced to admit that 'Smith's murderer' did not occur essentially and hence was not used

attributively. Thus synonymous expressions must be licensed substitutions for essentially occurring expressions. Further, there is textual evidence to suggest that Donnellan would accept the substitution of synonymous expressions, not merely under the duress of the above threatened counterexample, but freely. For what Donnellan considers important about an essentially occurring expression is not its phonetic or syntactic shape but its semantic force or meaning; it is not the words used to express an attribute but 'the attribute of being the so-and-so that is all important'. Since it is the attribute expressed that is all-important and not the words, there is no need artificially to restrict the permissible substitutions to synonymous expressions. Any expression of the same attribute, whether or not it is synonymous with the original expression, should be a permissible substitution; I shall call such expressions 'trivial variants'. Revising our explication we shall say, 'An expression occurs essentially just in case only trivial variants of that expression can do the same job'.

One remark remains to be made with respect to this explication of essential occurrences. Donnellan's phrase 'do the same job' is intuitive; yet it is perhaps clear enough. Within the standard theory of speech acts, two expressions do the same job just in case each would make the same contribution to a propositional and illocutionary act (see John Searle, *Speech Acts* (Cambridge, 1969), pp. 22—31).

With this explication of what an essential occurrence of an expression is, we can proceed with our example of a proper name used in such a way that it fits Donnellan's characterization of the attributive use. Suppose that a drawing has just taken place for the grand prize in some raffle for charity and the chairman of the raffle committee announces the winner by saying: 'Jane Smith has won the grand prize'. If we suppose that the chairman does not know anything more about Jane Smith, then there is no other name or description with which he can replace 'Jane Smith'. That is, the occurrence of 'Jane Smith' is essential, as required for the attributive use. Further, the chairman is using 'Jane Smith' in order to state that whoever is Jane Smith has won the grand prize. It does not matter whether Jane Smith completed the winning ticket, knew of the entry in her name or would have approved of it. (Of course, her winning can be voided if she refuses the prize.) Jane Smith, whoever she is, has won the grand prize. This latter point accords with Donnellan's further claim that if no one fits an expression used attributively, then the speech act is thwarted ('Reference and Definite Descriptions,' 291–292).

It might be objected that 'Jane Smith' is not essential because it can be replaced at least by the expression, 'The winner of the grand prize', which is not a trivial variant of 'Jane Smith'. The proper response to this objection is to explain why the proffered replacement is unacceptable. If we were to allow the replacement, then the chairman would say, 'The

winner of the grand prize has won the grand prize'. In such a case, he would fail to perform his intended speech act of announcing the winner because his new sentence, unlike the original, is uninformative. 'Jane Smith' contributes information that 'the winner of the grand prize' does not; thus the latter does not make the same contribution to the speech act and hence does not do the same job. Thus, the occurrence of 'Jane Smith' is essential.

It might also be urged that 'Jane Smith' does not occur essentially, because the description, 'The object named "Jane Smith"' can replace 'Jane Smith' and is not synonymous with the latter. (Depending upon one's other views, one or another reason might be given for asserting the non-synonymy. For example, if proper names have no meaning, they cannot have the same meaning as any description. Or, if all proper names are rigid designators and 'the object named "Jane Smith"' is not a rigid designator since Jane Smith might not have been so named, the two expressions will not always denote the same object in the same context, but synonymous expressions will.) In any case, whether or not 'The object named "Jane Smith"' is synonymous with 'Jane Smith', either the former is a trivial variant of the latter or it is not. If it is, then since it does the same job as 'Jane Smith', 'Jane Smith' none the less occurs essentially. If it is not, then by the same token no description occurs essentially; for every description can be replaced by a variant description, analogous to 'The object named "Jane Smith"', namely, the one formed by writing the original into the blank of the schema

$$\text{the object described by '}\text{\_\_\_\_\_}\text{'.}$$

And if no description occurs essentially, then none has an attributive use.

The results of this paper pose a dilemma for those who accept Donnellan's position. One horn of the dilemma is this: if there is an attributive use of definite descriptions and Donnellan has adequately characterized it, then there is also an attributive use of proper names; but this result is unacceptable to many philosophers, including Donnellan, who, as I mentioned earlier, denies that proper names have an attributive use. The other horn is this: if one denies that there is an attributive use of proper names, then either Donnellan has failed to characterize the attributive use adequately or there is no attributive use. The latter horn may be the easier one to take. For there are reasons, independent of one's views about the use of proper names, for rejecting all or part of Donnellan's position. One might claim that Donnellan's characterization of the attributive use is inadequate because of the vagueness of the essential/inessential distinction. Or one might claim that his characterization is inadequate because it fails to distinguish all referential from all attributive uses. Our example of the chairman's use of 'Jane Smith' is a case in point.

Not only is there no reason to think that the chairman is not using 'Jane Smith' referentially, but Jack Meiland, from whom I have taken this example, offers it as an example of referring in *The Nature of Intention*, pp. 48–9, though for a different purpose. Finally, if one becomes convinced that Donnellan's characterization of the attributive use is inadequate one might come to doubt the very existence of that use.

University of Texas at Austin   © A. P. MARTINICH 1977

AN ARGUMENT OF ARISTOTLE ON NON-CONTRADICTION

By H. W. NOONAN

WE are all familiar with the Fregean notions of sense and reference, and hence with the idea of two signs exhibiting sameness of sense and sameness of reference. However, if Saul Kripke's doctrine of necessary *a posteriori* truth is right there is another semantic equivalence relation which lies midway in strength between the traditional Fregean ones.[1] Let us say that two predicates '*F*' and '*G*' *have the same signification* if and only if it is necessarily true that $(x)(Fx \equiv Gx)$; and let us say similarly that two proper names '*a*' and '*b*' have the same signification if and only if it is necessarily true that $a = b$. On Fregean principles there can now be no objection to our countenancing as entities significations of predicates and proper names: for we have an exact statement of the criteria of identity for such entities. But, of course, this is *all* we know about these entities; hence any identification of them with some previously recognized entities which does not conflict with this knowledge is permissible. Thus we may identify the *signification* of a proper name with its *reference* if Kripke's view that identity statements containing proper names are all either false or necessarily true is correct, since in this case sameness of reference for proper names will both *entail* and be entailed by sameness of signification, as explained above. A similar identification cannot be made in the case of predicates of course, for not all coreferential, i.e. coextensive, predicates are *necessarily* coreferential. Thus in this case we need a mode of designation of the signification of a predicate which is distinct from our modes of designation of its reference and sense. I shall let '*to be (an) F*' (italicized) designate the signification of the predicate '*F*'. Thus *to be (an) F* will be *to be (a) G* just

---

[1] The existence of this middle strength equivalence relation *can* be recognized even if one does not accept Kripke's doctrine, since e.g. analytically equivalent predicates need not have the same sense—if sameness of sense entails substitutivity *salve veritate* in indirect speech contexts. It is just that the Kripkean doctrine makes its existence impossible to ignore.

in case it is necessarily true that $(x)(Fx \equiv Gx)$. Assuming that necessary propositions are necessarily necessary it follows that if it is true that *to be (an) F* is *to be (a) G* it is necessarily so. Thus designations of the form '*to be (an) F*' are like proper names (if Kripke is right about the latter) in that their significations may be identified with their referents. Since the reference of '*to be (an) F*' is, by definition, the signification of the predicate 'F' it follows that the signification of '*to be (an) F*' may be identified with the signification of 'F'.

I now want to use these Kripkean notions to make some suggestions about the interpretation of Aristotle's discussion of the Principle of Non-Contradiction (i.e. $\Box(x)(F) \sim (Fx \ \& \ \sim Fx)$) in Book Γ of the *Metaphysics*.

Aristotle says that this 'strongest of all principles' cannot be proved to someone who denies it, except by way of refutation—if he will only commit himself to 'signifying something'. He does not have to assert something to be or not to be; but only to signify something, both to himself and to someone else.

It seems that what he wants is the utterance of a significant name. For, he says, the first thing that is plain is that the name signifies to be or not to be this particular thing, so that it could not be that everything was so-and-so and not so-and-so. That is, it is still open to a man—so far as the argument has gone—to say that something can *both* be *and* not be this particular thing, but at least there is something definite that he would be saying it was and was not; so we are not in a state of complete flux and vagueness: not everything is a matter of 'so-and-so and not so-and-so' (a reference to Plato's *Theaetetus*).[1]

He goes on to suppose that the name uttered was 'man' and argues for the permissibility of his assuming that it 'signifies something and signifies one thing' (1006b12). And he then proceeds to argue that given this it is impossible that 'to be a man' should signify just what 'not to be a man' signifies (1006b13). This is the point at which commentators begin to stutter. In his notes to his translation of Books Γ, Δ, E Kirwan calls the opening clause of the paragraph beginning at 1006b13 'the first major crux' in the second argument for the PNC he distinguishes in Aristotle's text. The whole first sentence is the following (I give Kirwan's translation, see his *Aristotle's Metaphysics* Γ, Δ, E, Clarendon Press, Oxford, 1971):

> Then it is not possible that 'to be a man' should signify just what 'not to be a man' [signifies], if 'man' signifies not only about one thing but also one thing (for we do not count as signifying one thing this, viz. signifying about one thing, since in this way 'artistic' and 'pale' and 'man' would signify one thing, so that all will be one because synonymous).

[1] This interpretation follows G. E. M. Anscombe, *Three Philosophers*, 'Aristotle', Basil Blackwell, Oxford, 1967.

The problem is that there are conclusive objections against the two *prima facie* plausible candidates for what the first clause means. It might be taken as saying that 'is a man' and 'is not a man' cannot have the same meaning (sense), but that is a presupposition of the enquiry, as Aristotle himself is clearly aware, so why should it here be treated as something in need of argument?[1] Secondly, it might be taken as saying that 'is a man' and 'is not a man' have incompatible meanings (senses), i.e. cannot hold of the same thing. But if so it anticipates the conclusion of the argument from 1006b28–34 and consequently makes both the intervening material and that argument superfluous. Moreover, this interpretation of the first clause requires 'signifying one thing' to mean 'holding of one thing' which is a sense Aristotle explicitly reserves for 'signifying about one thing'. A third possible interpretation of the first clause, which Kirwan does not mention, is that it is saying that 'is a man' and 'is not a man' cannot have the same extension, but although this interpretation avoids the objections made to the other two, it is obviously inadequate: the point is that coextensiveness does not license substitutivity *salve veritate* in modal contexts, but at 1006b28 it is inferred from 'man' signifying two-footed animal that 'it is . . . necessary, if it is true of anything to say that it is a man, that it be a two-footed animal': so sameness of signification cannot just be sameness of extension. We see then why Miss Anscombe should say 'it begins to look as if [Aristotle's] sense of "signifying" is *none* of those currently in use among present day and recent philosophers' (*Three Philosophers*, p. 41).

I suggest that identifying Aristotle's notion of signification with our Kripkean notion of signification and taking his expressions 'to be a man' and 'not to be a man' to mean the same as our *'to be a man'* and *'to be a non-man'* avoids the various difficulties these three interpretations face. For, first, since sameness of signification does not entail sameness of sense, we may presuppose difference of sense while regarding difference of signification as something that needs to be shown. Second, since difference of extension is not guaranteed by difference of signification, a proof that predicates '*F*' and '*G*' do not have the same signification leaves it open that something may be *both F and G*. Third, that Aristotle regards sameness of signification as a sufficient condition of substitutivity *salve veritate* in modal contexts is no problem for this interpretation since our notion of signification is, of course, defined in such a way as to guarantee such substitutivity.

With the first clause of the quoted passage interpreted in this way we must interpret the statement that 'man' signifies one thing as meaning

---

[1] More exactly, it is a presupposition of the enquiry that 'is a man' and 'is not a man' have *different* meanings, but if one were to interpret the clause at 1006b13 as arguing that 'is a man' and 'is not a man' cannot mean the same one would likewise have to interpret the sentence at 1007a1 (given in translation below) as *arguing* that 'is a man' and 'is not a man' have different meanings.

that 'man' has a single signification (in our sense) whether or not it has
several senses. But now it becomes immediately obvious that the
impossibility of 'to be a man' (i.e. *'to be a man'*) and 'not to be a man' (i.e.
'to be a non-man') signifying the same cannot follow merely from *this*:
for suppose 'vixen' signifies a single thing—it will not follow that *to be a
vixen* cannot be *to be a female fox*. Aristotle's argument needs some further
premiss about the relation of 'is a man' to 'is a non-man' which is not
supplied at this point.

This conclusion agrees with Aquinas' understanding of the argu-
ment.[1] According to Aquinas the needed premiss is supplied at 1007a1:

> For 'to be a man' and 'to be a not-man' signify something different, if
> even being pale and being a man are different. For the former is much
> more strongly opposed so that it signifies something different.

On our interpretation this says that *'to be a man'* and *'to be a non-man'*
must signify different things if even *to be pale* and *to be a man* are different.
But why so? I think the answer is quite simple; we know from obser-
vation that not all men are also non-men and not all non-men are men,
just as we know from observation that not all men are pale and not all
pale things are men—in fact men who are non-men seem a good deal
scarcer on the ground than men who are pale ('the former is more
strongly opposed'). But then *'to be a man'* must signify something
different from *'to be a non-man'*, just as it must signify something different
from *'to be pale'*—and this holds without prejudice to the question
whether 'is a man' and 'is not a man' may not be verified of the same
subject. Hence, unless Aristotle's opponent wishes not only to deny the
PNC but also to maintain that $(x)(F)(Fx \,\&\, \sim Fx)$ he cannot regard
Aristotle's use of this premiss as a begging of the question. (And if he
does wish to maintain $(x)(F)(Fx \,\&\, \sim Fx)$, of course, Aristotle argues he
will be committed to all things being one.)

But with this premiss supplied the argument at 1006b13 goes through
(assuming that Aristotle's statement that 'man' signifies a single thing
is meant to cover also 'non-man''s signifying a single thing). For, in
general, for any predicates 'F' and 'G' if (1) each signifies a single thing
and (2) *'to be an F'* signifies something different from *'to be a G'* it follows
that *'to be an F'* cannot signify the same as *'to be a G'*. It is crucial to see
the difference between the second premiss of this argument and its
conclusion. The premiss is just that there are significations s and s' such
that $s \neq s'$, *'to be an F'* signifies s, and *'to be a G'* signifies s'. In the absence
of premiss (1) this is obviously compatible with there being a signification
s'' (possibly s or s') which both *'to be an F'* and *'to be a G'* signify: in
particular, the premiss stated at 1007a1 is compatible with the falsity of

[1] See St. Thomas Aquinas, *Commentary on the Metaphysics of Aristotle*, trans. J. P. Rowan,
Henry Regnery Company, Chicago, 1961, Vol. I, Lesson 7, Section 622 of Commentary.

the conclusion drawn at 1006b13 in the absence of the premiss that 'man' (and also 'non-man') signifies a single thing.

The relevance of the distinction Aristotle makes between 'signifying one thing' and 'signifying about one thing' should now be easy to appreciate: given that 'man' and 'non-man' each signify a single thing then if '*to be a man*' signifies something different from what '*to be a non-man*' signifies '*to be a man*' cannot signify the same as what '*to be a non-man*' signifies. But if 'man' merely signified *about* one thing, i.e. if it was merely the case that there was some object of which 'man' was true in all its various significations, then that '*to be a man*' and '*to be a non-man*' signified different things would obviously *not* rule out the possibility of their also signifying the same thing. Hence to make clear the force of his argument Aristotle emphasizes that 'signifying one thing' does not just mean 'signifying about one thing'.

What follows is relatively plain sailing for a short while. Aristotle now concludes that what 'man' signifies will not be to be and not to be the same thing (viz. a man) unless homonymously—that is, unless what we term 'man' in one language others term 'not-man' in another. But it is not this possibility, he points out, which his opponent wishes to maintain, but rather that something may both be and not be a man in actual fact, that is, that 'man' and 'not-man', in the different meanings they have in *our* language, may be true of the same subject. ·

The next major problem of interpretation is the section from b22–8. As Kirwan says, there are two main problems here. First of all there is the problem of understanding the purpose this section is intended to have. And secondly there is the problem of understanding exactly how the argument as it is presented in the text is meant to go. It is clear enough that the argument of the section is that 'man' and 'not-man' must signify something different given that, as has been shown, 'to be a man' and 'not to be a man' do so. But first, why does Aristotle wish to argue this, and second, how exactly *does* he argue it? I think Miss Anscombe gives the right answer to the first of these questions when she describes the section as arguing that if 'being A' does not signify what 'not being A' signifies, then 'A' *as a predicate* cannot signify what 'not A' *as a predicate* signifies (*Three Philosophers*, p. 41). The point is that Aristotle's argument so far has only been about the significations of 'to be a man' and 'not to be a man'—which are his own technicalities. But his opponent, of course, will not use these technicalities in stating his opposition to the PNC—he will use ordinary Greek. But then in the absence of the argument of b22–8 he could claim that Aristotle's discussion had no relevance to his position.

Our second problem was to understand the exact structure of b22–8. The difficulty here is that given that the argument intended is the one described above, the presentation of it in the text introduces irrelevant

material—as becomes perfectly clear once it is interpreted by means of our Kripkean notions. So interpreted the argument begins: if 'man' and 'not-man' do not signify something different then neither do *'to be a non-man'* and *'to be a man'*—*already* then Aristotle is in a position to deduce that 'man' and 'not-man' signify something different by appealing to the previously established conclusion about *'to be a man'* and *'to be a non-man'*; but instead he goes on irrelevantly: so that *to be a man* will be *to be a non-man*, for they will be one thing (for to be one thing signifies this: being like mantle and cloak if the formula is one); and then argues round in a circle: but if they are one thing *'to be a man'* and *'to be a non-man'* signify one thing; and only *now* points out that he has already shown this to be false.

But now: in what sense does this meandering course pose a difficulty of *interpretation*? We have all had the experience of presenting a basically valid argument in an imperfect and repetitious form (at least, I have), so why should we not just say that this is what Aristotle has done here? There can be no doubt, after all, about the basic argument intended or (if we interpret it as I suggest) of its validity, so unless we expect invariable logical perfection from Aristotle I cannot see that any puzzle remains.

With this I reach 1006b28. The argument which begins at this point presents a good many puzzles, e.g. there is the question of whether the modality should be taken as *de dicto* or *de re*, and hence whether Aristotle should be regarded as defending the unrestricted PNC or a version whose scope is restricted to substantial (or essential) predicates. But I shall finish by discussing just *one* of these puzzles, namely, what use does the argument here make of the results Aristotle has previously established? It is plain enough, and especially so in the light of our Kripkean interpretation, how Aristotle's argument depends upon the assumption that 'man' signifies two-footed animal and nothing other than two-footed animal—this is so whether we take the modality *de dicto* or *de re*. But it is equally clear, in the light of this interpretation, that it does *not* make any use of the point that 'man' and 'not-man' cannot signify the same thing. So what was the purpose of Aristotle's previous demonstration of this? I argued earlier that he had to show how it followed from the impossibility of 'to be a man' and 'not to be a man' signifying the same that 'man' and 'not-man' *as predicates* could not signify the same in order to bring his discussion to bear on his opponent's position, but I did not then say *what* relevance his demonstration that 'to be a man' and 'not to be a man' could not signify the same was intended to have. This is the question now confronting us. I think Aquinas' suggestion must be right (*Commentary*, Section 621): Aristotle has argued that 'man' and 'not-man' cannot signify the same thing, not so much in order to further his own argument for the PNC but to forestall a possible counter-argument from his opponent. For suppose 'man' and 'not-man' did

signify the same; then it would be as valid to infer that whatever was a man necessarily was *not* a two-footed animal as to infer that whatever was a man necessarily *was* a two-footed animal, i.e. it would be *no more* valid to make the latter inference than to make the former. But, of course, Aristotle cannot hold that both these inferences are valid, so long as he maintains the PNC, so he must say either that neither is or that 'man' and 'not-man' do not signify the same. But, then, in the absence of proof that 'man' and 'not-man' do not signify the same, his opponent could say that he had no right to assume that *either* was valid—given his acceptance of the PNC. And so Aristotle provides such a proof.[1] As Aquinas puts it:

> Now the things demonstrated above are useful to his thesis, because if someone were to think that the terms man and not-man might signify the same thing, or that the term man might signify both being a man and not being a man, his opponent could deny the proposition that man must be a two-footed animal. For he could say that it is no more necessary to say that man must be a two-footed animal than to say that he is not a two-footed animal, granted that the terms man and not-man signify the same thing, or granted that the term man signifies both of these—being a man and not being a man.

[1] In opposing the PNC Aristotle's opponent is not denying that some of the things Aristotle believes are true, but rather is claiming that *more* things may be true than Aristotle believes to be possible, viz. the contradictories of some of the true propositions which Aristotle believes. Thus it is natural for him to confront any argument Aristotle puts forward, not by denying its validity, but by proposing another argument whose conclusion is the contradictory of the conclusion of Aristotle's argument, and then challenging Aristotle to explain why just *his* argument is valid. I take the purpose of the paragraph beginning at 1006b13 to be to forestall just such a challenge.

Trinity Hall, Cambridge   © H. W. NOONAN 1977

FURTHER NOTES ON FUNCTIONS

By PATRICK GRIM

MY beloved stuffed moose head is functioning as a hat rack, and occasionally functions to hold coats and umbrellas. But it is really a decorative, rather than functional, addition to my study; it is the function of my hat rack to hold hats, and a function of the coat rack to hold umbrellas. The moose head often functions only to frighten small children.

Attention to the many different ways in which we speak of functions is called for by the flurry of recent work done on the topic, and especially by Christopher Boorse's recent contribution (*Philosophical Review*,

January, 1976). Boorse's article involves a critique of Larry Wright's original formulation for functions (*Philosophical Review*, April, 1973), using in part counter-examples similar to my own (ANALYSIS 35.2, 1974). Wright seems to recognize the force of such an attack (ANALYSIS 36.3, 1976).

As a critical piece, Boorse's article has much in its favour. Boorse also attempts, however, to supply what Wright as of this writing has not: an adequate account of precisely what it is that talk of functions amounts to. In the end, I think, Boorse's proposal is less satisfactory than the account it is designed to replace.

Boorse claims that 'functions are, purely and simply, contributions to goals' (p. 77), and offers an analysis of 'X is performing the function Z . . .' on the basis of an outline of 'goals'. He goes on to discuss 'the function of X is Z' and 'a function of X is Z' in terms of 'X is performing the function Z . . .' The three main steps in his presentation are thus: (1) the outline of 'goals', (2) the analysis of 'X is performing the function Z . . .' in terms of 'goals', and (3) the transition from 'X is performing the function Z . . .' to 'the function of X is Z' and 'a function of X is Z'. I hope to point out important difficulties which arise with each step.

I

Boorse first specifies 'goals', following Sommerhoff in *Analytical Biology*, as follows

> To say that an action or process A is directed to the goal G is to say not only that A is what is required for G, but also that within some range of environmental variation A *would have been modified* in whatever way was required for G (p. 78).

Boorse discusses two objections to such an account proposed by Scheffler (*British Journal for the Philosophy of Science*, 9, 1959), and alters the outline of 'goals' in terms of them. The first objection Boorse considers and attempts to resolve as follows

> Presumably a cat which waits by an empty mousehole may have the goal of catching a mouse; but it is hard to see how any behavior can literally be required for catching a nonexistent mouse. The cat's behavior can, however, fairly be called appropriate to catching a mouse; it is, for instance, the kind of behavior that leads to catching mice when they are there. And this answer seems sufficient . . . (p. 79).

What Boorse is proposing, I think, is the following amendment of the outline of 'goals' above:

> To say that an action or process A is directed to the goal G is to say not only that A is what is required for G *or appropriate to G-ing*, but also that within some range of environmental variation A *would have been modified* in whatever way was required for G.

Cat and mouse problems, and perhaps even Scheffler's original cat and mouse problems, remain. Suppose that kitty lurks around empty mouseholes with the goal of catching mice. Not only does it appear that lurking around empty mouseholes is not required for catching mice, it also appears that lurking around empty mouseholes is totally *inappropriate* to catching mice. Nor, in the end, is 'lurking around empty mouseholes' the kind of behaviour which leads to catching mice when they are there, since if there are mice in the hole kitty's 'behaviour' can no longer be described as 'lurking around empty mouseholes'. Lurking around mouseholes may lead to catching mice when they are there, but no logically respectable cat could possibly lurk around empty mouseholes full of mice.

We can, of course, make things still worse for the formulation. Poor kitty, demented as she is, commonly does totally unrequired and inappropriate things with the goal of catching mice. If Boorse's formulation for 'goals' were correct, we would be forced to say that doing totally unrequired and inappropriate things is here required for or appropriate to catching mice.

Boorse also considers a second difficulty, and once again revises the account of 'goals' in terms of it. So as to avoid difficulties with 'behaviours' required for or appropriate to several ends, Boorse specifies that

> When a process appropriate to several ends at once has a true goal, I suggest it is because the process is produced by an internal mechanism which standardly guides pursuit of that goal but not the others (p. 79).

With that amendment, the entire formulation becomes:

> To say that an action or process A is directed to the goal G is to say not only that A is what is required for G, *or appropriate to G-ing and if appropriate also to any other X-ing is produced by an internal mechanism which standardly guides pursuit of G but not of any other X*, but also that within some range of environmental variation A *would have been modified* in whatever way was required for G.

The attempt to avoid one type of difficulty here, I think, quickly raises another. Thus consider the following case. Kitty's one central goal in life is to appear sinister, and everything she does is guided by that goal in one way or another. She often crouches as if to pounce, silently glaring and extending her claws, solely in order to appear sinister. Should some additional need to catch mice or annoy her mistress arise, kitty's central goal of appearing sinister will guide these efforts as well; she catches mice and annoys her mistress in as sinister a manner as possible. If kitty's central goal of appearing sinister has a 'mechanism' m

(little electrical trails in kitty's brain, perhaps), m is part of the 'mechanism' of everything kitty does.

Kitty is at present crouching near the mousehole, silently glaring and extending her claws, and is doing so merely with the goal of appearing sinister. Such behaviour is also appropriate to catching mice or annoying her mistress, of course, but kitty's goal in this case is merely the familiar one of appearing sinister. Here the formulation above would seem to require that kitty's present 'behaviour' be 'produced by an internal mechanism which standardly guides' the attempt to appear sinister but does *not* standardly guide *other* attempts such as catching mice or annoying her mistress. But this is not the case; kitty's goal here, as well as whatever 'mechanism' goes with it, are such that they guide everything kitty does and thus do not guide merely the kind of thing she is doing at the moment. As it stands, the formulation above seems to exclude all such cases. If Boorse's account were correct, in fact, neither kitty nor anyone else could act solely in pursuit of a goal which also standardly guided the pursuit of other goals.

There finally seems to be some difficulty in the 'would have been modified' clause of the formulation. Certainly I might have a goal G which I know can only be accomplished by doing some very simple A or some very strenuous A'. I do A with the goal of G, but had A failed to produce G I would have totally abandoned G rather than subject myself to the rigours of A'. Here action in terms of a goal seems obvious, though the specifications of the formulation are not fulfilled in that my action would not have been modified within any 'range of environmental variation . . . in whatever way was required for G'.

Boorse's attempt is not simply to clarify ordinary notions of goals, since he wants to label thermostats and guided missiles as 'goal-directed' as well and speaks of 'goal-directed behaviour outside of the realm of intentional action' as something like 'a theoretical concept of biology to be explicated according to convenience' (p. 78). To that extent, his earlier claim that 'functions are, purely and simply, contributions to goals' (p. 77) becomes misleading without inverted commas around 'goals' and a clarificatory footnote. But Boorse also seems to want his formulation to present necessary conditions for 'goals' in the ordinary sense, or at least to avoid difficulties of a type he considers, and with an eye to either of these standards he appears to have failed on three counts.

II

For the sake of argument, however, we might allow Boorse's discussion of goals as a discussion of 'goals' in a particular sense, and consider his definitional efforts as purely stipulative ones. All of this would be harmless if the analysis of 'functions' came out right.

Boorse offers the following as an account of 'what is perhaps the weakest of all functional attributions'

> X is performing the function Z in the G-ing of S at t, *means*
> At t, X is Z-ing and the Z-ing of X is making a causal contribution to the goal G of the goal-directed system S (p. 80).

'G' and 't' here are variables for goal and time. 'S' stands for 'system', a term for which we are never really given an explanation.

I must confess to some difficulty with the phrase 'is performing the function Z' simply because it sounds so pretentious (much as 'believe' commonly sounds in speaking of what people think and as 'performing the action a' commonly does in speaking of what people do). I hope my words are functioning to get my message across, but to say that I hoped they were 'performing the function of getting my message across' would seem at least unnecessarily awkward. Boorse, in fact, never uses precisely that phrase in context. 'Is functioning to' seems a more natural candidate for the 'weakest of all functional attributions', but I don't want to saddle Boorse with a phrase he did not attempt to analyse.

Neither phrase, however, seems right when we are speaking of people in certain situations, and for that reason Boorse's account faces a simple form of counter-example. Thus consider a Rube Goldberg contraption the goal of which is to produce pretty patterns on an oscilloscope. Rube has set the whole thing up in such a way that it relies in part on the snoring of Professor Emeritus, who knows nothing of the device and is lost to the world. In such a case it may be that Professor Emeritus is snoring and his snoring is making a causal contribution to the device's production of pretty patterns on the oscilloscope. But it would seem odd to say either that Professor Emeritus is performing the function of snoring, or that Professor Emeritus is functioning to snore. We can make the case worse by building a device which relies on Professor Emeritus lying there unconscious or even dead (we want a flat EEG reading, perhaps). Though he may be lying there unconscious or dead and that may contribute to the goals of the 'system', it would be very strange to claim that Professor Emeritus is either functioning to lie there unconscious or is performing the function of lying there dead.

Other difficulties face Boorse's account as well. Though he speaks of it constantly as an analysis of 'performing the function' and even at one point as an analysis of 'performing a function', his account is strictly speaking an analysis rather of the complex phrase 'performing the function Z in the G-ing of S at t'. As such, even if the account were successful, it would not follow that 'functions are, purely and simply, contributions to goals' (p. 77) nor that 'to accept our analysis of performing a function is to settle the question of what sort of thing a

function is—namely, a contribution to a goal' (p. 81). The phrase for which he presents an analysis is one which requires the mention of some particular 'goal' (as a substitution for 'G-ing') and thus the claim that such a phrase can only be analysed in terms of some 'goal' is a foregone conclusion and a fairly uninteresting one. The least that is required to substantiate Boorse's more general claims that 'functions are, purely and simply, contributions to goals' is an analysis in terms of 'goals' of some subspecies of function talk which need not explicitly mention 'goals'.

Boorse may have intended his analysis to be more than it in fact is; the account is followed immediately by the claim that 'all functional statements, weak and strong, seem to me implicitly relative to system, goal, and time' (p. 80), and he speaks later of 'variables suppressed' in 'function statements' (p. 85). If Boorse intended his account to cover cases where 'goals' are not explicitly mentioned, the following examples suggest direct counter-examples in terms of both 'performing functions' and 'functioning to'. Boorse's specific analysis aside, such examples make implausible any account of the form required to substantiate his more general claims regarding the necessity of an analysis in terms of 'goals'. Thus consider:

(1) The advertising department has for twenty years been performing a function detrimental to all goals of the company, society at large, and even the department itself.

(2) The seniority committee is to this day performing a function inimical to all goals of the legislative branch and government in general. (Give it a new name and it would simply be the same committee performing the same function inimical to the same goals.)

(3) The vast distances of space are functioning to prevent communication between us and our nearest star.

(4) The force of the current is functioning to sweep small boats quickly out of little boys' reach.

Each of these would appear to resist analysis in terms of 'goals', (1) and (2) because they explicitly deny that the work of the advertising department or the seniority committee furthers any goals which might seem relevant in such an analysis, and (3) and (4) because there seems to be no 'goal' in sight appropriate to such an account. Only if the waters of the world were in conspiracy against little boys would (4) fit such an account, and only if the universe were plotting against interstellar communication would (3) go through. There need be no such goals in any of these cases in either the common sense of the term or in that uncommon sense Boorse has defined. It was required of the latter that

'within some range of environmental variation *A would have been modified* in whatever way was required for *G*' (p. 78). But none of these cases need be ones in which anything would have changed. Departments and committees remain out of apathy, and rivers don't change their courses suitably to the loss of small boats.

III

Boorse's third attempt is to move from an account of 'performing the function *Z*' to a broader outline of 'the function of *X*' and 'a function of *X*'. Like its two predecessors, I think, this step in Boorse's discussion faces its own peculiar difficulties.

> Whatever 'the' or 'a' function of a thing is, then, it must at least be a contribution to a goal . . . what more is required for a function performed by *X* to be among 'the functions' of *X* is not any fixed general property but instead varies from context to context (p. 81).

Boorse's final accounts read:

> 'A function of *X* is *Z*' means that in some contextually definite goal-directed system *S*, during some contextually definite time interval *t*, the *Z*-ing of *X* falls within some contextually circumscribed class of functions being performed by *X* during *t*—that is, causal contributions to a goal *G* of *S*.

> 'The function of *X* is *Z*' means that in some contextually definite system *S* with contextually definite goal(s) *G*, during some contextually definite time interval *t*, the *Z*-ing of *X* is the sole member of a contextually circumscribed class of functions being performed during *t* by *X* in the *G*-ing of *X*—that is, causal contributions to *G* (p. 82).

These are not, Boorse warns, to be taken as analyses 'in the sense of two-place synonymy relations', precisely because of contextual variance. A very major problem arises with respect to the general attempt here nonetheless.

In his original article, Wright notes that

> The function of that button on the dashboard is to activate the windshield washer, even if all it does is make the mess on the windshield worse . . . That would be its *function* even if I never took my car out of the garage—or broke the windshield . . . If the windshield wiper comes from the factory defective, and is never repaired, we would still say that its *function* is to activate the washer system . . . (p. 146).

Boorse rejects this general claim out of hand as a 'curious ruling' (p. 73), but in fact Wright was correct first time around and this very point indicates an essential problem with Boorse's attempt to derive 'a function' and 'the function' from 'functions being performed'.[1]

---

[1] Strangely enough, Boorse considers this problem briefly in a footnote (p. 83) which itself seems to go directly against his explicit accounts.

Some things have a particular function, and the function they have is of a certain nature, despite the fact that there is no time t at which or during which they are performing that function, functioning, or even functional. The function of the light switch in the west bedroom is to turn on that light. But the electrician hooked it up wrong. It doesn't, and cannot perform any useful function and will probably never be functional. The function of a crab's claws is to ward off aggressors. That's true even of Charlie the crab's claws, though Charlie lives alone in my aquarium and will never meet any aggressor; there will be no time t at which Charlie's claws are functioning to ward off an aggressor. The function of Snowball's ears is to allow her to hear things, despite the fact that she is a white blue-eyed cat and there is nothing known to science which would allow the poor dear to hear things.

If Boorse's account were correct, moreover, we could never decide whether something had a particular function without knowing whether it ever would so function. A perfectly functional windshield wiper switch might never be used, or the entire car might explode before anyone gets close to the switch. But we need not wait and see whether any such thing happens before declaring its function to be the wiping of windows.

Thus the attempt to get 'a function' and 'the function' from 'is performing the function' seems doomed, as doomed as the attempt to define 'functioning' or 'is performing the function' in terms of 'goals' and as doomed as Boorse's outline of 'goals' itself.

IV

The notion of 'functions' and the attempt to give a clear account of it are important enough for all contributions to be applauded, and Boorse's work is important as a critique of accounts which have preceded it. For a more adequate account than those he criticizes, however, we must look elsewhere.[1]

[1] I am much indebted to Ms Kriste Taylor for the loan of her linguistic ear.

SAYING OF

By Jennifer Hornsby

I

THEY are variables. That is what people say about pronouns to signal their connection with the symbols appropriate to their formal representation. If all utterances in natural language were of closed sentences, then the right thing to say about pronouns would be 'they are bound variables'. But I think that we sometimes utter open sentences. This paper started with one, for instance; and the first word there—the pronoun 'they'—should be treated as a free variable.[1]

II

In 'On Saying That' ([5]) Davidson gave an account of the logical form of sentences of indirect discourse. He dealt only with what Quine has called *notional* readings of such sentences (v.[13]). It has been thought that his theory cannot manage *relational* saying, and that as a result problems about quantification into contexts of *oratio obliqua* remain untouched (v. e.g. [1] p. 289 and [7] p. 10). But if I am right in claiming that we utter open sentences, there is a natural way to extend Davidson's account. I propose such an extension below. I try to show, while I am at it, that the resulting account can be used to cover a whole range of attitudinatives, not just 'say'.

III

> Sentences of indirect discourse . . . consist of an expression referring to a speaker, the 2-place predicate 'said' and a demonstrative referring to an utterance.   Davidson [5] p. 170.

When Davidson wrote this he didn't concern himself with the fact that sentences of indirect discourse are *action* sentences and fall inside the scope of another of his own doctrines (v. [6]). If we say what someone said then we tell of a speech act that he performed, and we can if we want say when or where or how he did it. 'In 1609 Galileo said that the earth moves'. So we need to combine Davidson's account of the logical form of action sentences with his treatment of indirect discourse. If we do, we arrive at the following for this adverbially modified action sentence:

$(\exists u)(\text{In}(1609,u)$ & Said $(\text{Galileo},u,\text{that}))$. The earth moves.

---

[1] Note that I didn't use my first sentence to make an assertion. I used it rather to tell you what people say about pronouns. Suppose I had myself been talking about pronouns and used that same string to express my thoughts about them. Let me do so now. They are variables. Then again it is plausible that I have uttered an open sentence. But this time the free variable ('they') is assigned a determinate value in context conferring a determinate truth-value on my sentence (cp IV). Whereas the open sentence which started this paper was not so much as a candidate for truth or falsity.

The 'said' here is the tensed 'say' of indirect discourse which we actually use. Some people have found this sense of 'say' mysterious.[1] It is a sense in which I can tell you that another, far removed from us, *said* something which I refer to. So he, the person reported *said* what I (or you, my interlocutor) give voice to. The mystery arises when this indirect 'say' is confused with a direct 'say', henceforth 'utter'. The mystery is dispelled when we see how the two are connected:

Say(a,u,v) iff Utter(a,u) & SAMESAY(u,v).

So if I announce 'Tim says that it's raining' then I announce that an utterance of Tim's SAMESAYS with my utterance just past of 'it's raining', but not, of course, that Tim *uttered* my utterance.

Davidson introduced 'samesay' as a relation between *speakers* to explain the indirect 'say' (v. [5]). A relation like my 'SAMESAY', which holds between (a) states or events (which may be actual utterances) and (b) utterances, has the promise of elucidating a wider range of notions. For instance, it is an attractive suggestion that all statements of propositional attitude (with "that-clauses" appended) have logical forms on the same model.[2] So, e.g.:

($\exists z$)(Fears(A,z,that). It's raining

($\exists z$)(Believes(B,z,that). It's raining.

And a meaning analysis would show that these imply that a "fear state" of A and a "belief state" of B SAMESAY with the respective utterances of 'it's raining'.[3]

IV

So much for a view about notional cases. The announced problem was the relational sentences. Consider

(1)   Galileo said of the earth that it moves.

---

[1] E.g. McFeteridge in [11]. He points out that on Davidson's account the number of things which Galileo said ought to be the number of token utterances referred to by 'that's in utterances of the form 'Galileo said that'. But there might have been (e.g.) ten thousand utterances of 'Galileo said that the earth moves' even though Galileo said 'Eppur si muove' (or words to that effect) five times. On the present view there is room to suppose that it is Galileo's direct sayings whose number we should ask after if we were to enquire 'How many things did Galileo say?', i.e. that we should ask how many verifiers or 'u', not 'v', in '($\exists u$)($\exists v$)(Said(Galileo,u,v))' there are. I don't deny that problems remain about connecting the required direct 'say' with the 'say' of indirect speech.
[2] Schiffer gives grounds for the view for the special case of 'desire' ([15] p. 199–200). The grounds apply to the cases of 'believe' and 'hope' and 'fear'.
[3] There is no precisely parallel analysis in these cases because there is no 2-place predicate which could serve in the case of 'believe' or 'fear' to replace 'utter' as it occurs in the analysis of 'say'. This will only worry those who make the mistake of identifying the analysis with the logical form.

My suggestion for sentences like (1) falls out of the remarks in I and III above. It is this. The relational (1) differs from its notional counterpart—'Galileo said that the earth moves'—in just two simple ways.

(a) (1) contains an adverbial phrase, a conjunct in logical form which says what Galileo's utterance was 'of', what it was about.

(b) In (1) the demonstrative 'that' refers to an utterance of an open sentence—the 'it' in the utterance it picks out corresponds to a free variable.

(Support for the view that the 'it' in (1) is a free variable is derived from its explanatory power. But it is not bound by any quantifier, for 'the earth' is a singular term so that there is none for it to be bound by;[1] and it is not a pronoun coreferential with 'the earth', for if it were then reference to the earth would be reintroduced after 'that', so that 'the earth' could not be expected to occur, as it does, wholly transparently.[2]) We have, then:

(1*) $(\exists u)(\text{Of}(\text{the earth},u)$ & $\text{Said}(\text{Galileo},u,\text{that}))$. x moves.

The only new primitive here is 'of'; the 'say' we found before. So, if the analysis suggested in III is right, then, provided (1) is true, an utterance of Galileo's stands in the *SAMESAY* relation to my utterance of 'it moves'. This may seem problematic, for though I uttered an open sentence and one which expressed no complete thought, Galileo expressed a quite determinate thought when he said 'Eppur si muove'. But there need be no problem here, even if we take Galileo's own sentence to have been a closed sentence. We should then have to let 'SAME-SAY' relate utterances of closed and open sentences. But we could say, quite generally, that a closed utterance u SAMESAYS with an open utterance v if and only if u is of an object and what it predicates of that object is the same in import as v. That would be one view.

Tyler Burge has argued, however, that sentences containing demonstratives or proper names (which he argues are composed from demonstrative expressions, v. [4]) are open sentences, albeit that utterances of them may be determinately true or false. If this is the right account of relational assertions (cp. [17] p. 93), no wonder that open sentences can be used to give their content. Galileo's famous words

[1] If you doubt that 'the earth' is a singular term, then pick another example.
[2] This isn't to say that there are no sentences in which 'the earth' behaves both transparently and opaquely; indeed I agree with Loar (v. [10]) that terms in surface English very often have this dual role. Nor is it to say that a word's occurring after 'that' in English guarantees its opacity. My point applies to (1)—a sentence of logician's English—heard with a logician's ear. (I think superficially notional sentences in English have relational readings which combine features of the pure notional with features of "logician's" sentences. Cp. n.9.)

provide a good example: in context he secured reference to the earth, but 'Eppur si muove' contains no explicit mention of the earth. Once we have said that it was the earth of which he spoke, our open sentence suffices to say what he said.

Whether or not one adopts Burge's view, it is a consequence of my account that 'it moves' could be used to report any utterance about any object which was thereby said to move. But there is no harm in this. The new found looseness in what SAMESAYS with a given utterance is in practice always compensated by further predications of the utterance; one must tell what it was *of*.

The account can be used straightforwardly to deal with 'fears of', 'believes of' and the rest. And now the ontology of states does some work. If Tim has a belief of the Queen, then it is *via* his belief state that he is related to the Queen. To say more about the 'of' relation here would be to start on an account of what it is for a mental state to be *de re*. But my task here is the more limited one of uncovering the logical form.[1]

V

(2) $(\exists u)(\exists y)(\text{Of}(y,u)$ & $\text{Said}(\text{Ralph},u,\text{that}))$. x is a spy.
(3) $(\exists u)(\text{Said}(\text{Ralph},u,\text{that})$. $(\exists x)(x$ is a spy.)

These are the logical forms of 'Ralph said of someone that he's a spy' and 'Ralph said that there are spies'. There is no quantifying in—the scope of '$(\exists y)$' in (2) ends with the full stop. Indeed the departures from Quine are slight.[2] But the polyadicity of 'say' is invariable, and no quotation is introduced when the 'that's are treated Davidson's way.

There is no entailment, in either direction, between (2) and (3) in virtue of their form. And the present example of Ralph's sayings suggest that this is the right result. But in other cases there are inferences which we should want to explain. E.g. from 'Ralph said of Dick that he's a spy' to 'Ralph said that Dick's a spy', and, switching to *belief*, from 'Ralph believes of someone that he's a spy' to 'Ralph believes that there are spies'. The first of these is registered in the meaning analysis granted natural assumptions about 'SAMESAY' and 'of'; the second requires for its explanation a further principle to the effect that a person believes the existential generalization of anything he believes. In general, genuine inference of this kind will be warranted by specific principles for

---

[1] See Kaplan ([9]) for an account, though Kaplan took himself to be giving logical forms.

[2] In [14] p. 335 Quine even suggested extending Davidson's account to relational sayings . And Temin [16] has proposed finding reference to utterances of open sentences in 'saying of' sentences. But neither of these sees the unification of 'says that' and 'says of that' proceeding quite as I do.

particular attitudinatives. From these we should hope to predict such inferences as particular contexts sustain.[1]

VI

The examples so far considered have all featured only 1-place open sentences, so that it has been in the relational cases just a plain individual related by 'of' to utterances (for 'says', a state for 'believes', 'hopes' or 'fears'). And it is one thing to give a unified account of just this simple kind of relational saying and the notional: it is another to show that that account will work for more complex relational sentences, those of higher degree. We must treat such sentences as these

(4)   Ralph says of Dick and Tom that the former spies on the latter.

(5)   Tim believes of Oxford, Cambridge and London that the first lies between the second and the third.

What should one say about them?

One could introduce new 'Of' relations—a 3-place one for (4), a 4-place one for (5). But those who were troubled with previous accounts precisely for crediting 'say' with multiple ambiguity would scarcely be satisfied with an account which required indefinitely many 'Of' relations. It is small comfort to learn that 'say' remains univocal in combination with them all. We do better to suppose that 'Of' is constantly 2-place and that they are *pairs* and *triples* that (4) and (5) say utterances were of. Designations of *n*-tuples (to requisitely high *n*) can be introduced as primitives.

I can imagine objections to the new object-language ontology. But its introduction commits us to no more than the supposition that speakers have the concept of an order. And we know that English speakers know that there is a difference between (4) and

(6)   Ralph says of *Tom and Dick* that the former spies on the latter.

We recapitulate their knowledge by distinguishing the pair denoted by 'Tom and Dick' from the pair denoted by 'Dick and Tom'.

Thomas Baldwin has claimed that we cannot make all the distinctions necessary here without introducing *sequences*, conceived as logicians conceive them, into the object-language. And this is a step which he is not prepared to take.[2] Only if we identify the things talked about in 'saying of' statements with (*inter alia*) such sequences, Baldwin thinks,

[1] I discuss some of the issues which arise here in [8]. See also Wallace [17]. I believe Loar's view in [10] to be crucial to explaining all our intuitions about inferences of this kind. But I shan't go into the matter here.

[2] In [2], I have had to restate Baldwin's objection to apply it to my own account; but I take it from what he says that he would press this point against me.

can we appropriately connect up the order in which individuals are stated to be related to sayings by 'of' with the order in which they are stated to have been said to be related by ' . . . spies on . . .'. Now it is true that we have to make out a difference not only between (4) and (6) but also between (4) and (7):

(7)   Ralph says of Dick and Tom that *the latter* spies on *the former*.

But what this shows is that English speakers manifest their possession of the concept of an order twice over in sentences like these: they distinguish between 'Tom and Dick' and 'Dick and Tom' *and* they distinguish between 'the former spies on the latter' and 'the latter spies on the former'. All that the theorist need do is capture both sorts of distinction in his representations of the sentences and let it be shown, in the semantics, that each corresponds to a genuine difference. And there will be no confusing (4) with (7) so long as it is made clear that the utterances which are demonstrated by the 'that's which occur in each of these are utterances of different type sentences.

One way to distinguish between the open sentence in (4) and (7) is to distinguish between 'x_1 spies on x_2' and 'x_2 spies on x_1', to which there corresponds a genuine semantic difference since different sequences satisfy each of these. Let it not be complained that these indexed variables are merely an "ingenious technical device" corresponding to nothing which speakers understand. The use of variables here is no more a technical one than their standard use to replace pronouns bound by quantifiers; and the use of indices is no more ingenious than the speaker's use of expressions like 'the former' and 'the latter' or 'the first', 'the second' and 'the third'. Speakers display their concept of an order, the theorist simply makes it manifest with the tools at his disposal.

VII

Here is (4) in a form fit for semantic treatment:

(4*) $(\exists u)(\mathrm{Of}(<\mathrm{Dick,Tom}>,u)$ & $\mathrm{Said(Ralph},u,\mathrm{that}))$. x_1 spies on x_2.

The captive sentence is unproblematic.[1] I show that we can deal with the "saying of" sentence.

'of' and 'said' are 2- and 3-place predicates and receive standard treatment. ('u' can be replaced by a suitable ("ordered") variable.) '$<$Dick AND Tom$>$' is a singular term. We say $s*('<$Dick AND Tom$>')$ =

---

[1] Of course we can only give satisfaction conditions for the open sentence. But there is no special need here to credit speakers with explicit knowledge of them.

$<s^*(\text{'Dick'})$ AND $s^*(\text{'Tom'})>$;[1] (Generally $s^{*\lceil} <t_1,t_2 \ldots, t_{n-1}$ AND $t_n>^\rceil = <s^*(t_1), s^*(t_2) \ldots, s^*(t_{n-1})$ AND $s^*(t_n)>.)$[2]

The demonstrative 'that' is treated as a free variable (cp. Burge [3])—say it is x_0. Following Burge: for all speakers p, times t, if p uses 'x_0' (in a sentence from a syntactically specified class) at t to refer TO an utterance z, then for all sequences s, $s^*(x_0) = z$; otherwise $s^*(x_0) = $ the 0th member of s.

A theory of reference will tell us what it is for p to refer at t TO an utterance. But we can safely say that it will be a necessary condition of an utterance being referred to that that utterance be by p or his audience and at a time close to t, and that if any token 'that' refers to any utterance then it refers to just one such. If no utterance is referred to then the sentence takes on no determinate truth-value since different objects will be assigned to 'that' ('x_0') for different sequences, making some sequences satisfiers of 'Said(Ralph,u,that)', some not. But we have still shown what truth for the sentence *would* have consisted in if only 'that' had referred.

VIII

Understanding speech about particulars requires a grasp of correlations between utterances and objects in the world. Accurate reporting of speech about particulars requires mention of the objects correlated with particular speech acts. So a view about ordinary discourse lends support to the account of *oratio obliqua*.

It is tempting for the theorist to exploit this connection between direct and indirect speech. In his theory of ordinary speech, correlations are set up between utterances and objects: objects are values of sequences and it is sequences that utterances are true of. So why not suppose that when speech is reported in natural language it is just these *sequences* the person reported is said to have spoken of?

This is the suggestion that we read the ontology of our theory into the object-language: the speaker who says what another has said uses just so much theory.[3] The suggestion brings no errors about the form of speech reports; but there is no actual evidence that it correctly identifies their content. And the fact that we are able to treat relational *oratio obliqua* without supposing that sequences (i.e. functions from natural numbers to individuals) are the content of object-language talk provides the best possible reason for thinking that speakers need not here be

[1] For clarity I write the axiom as if proper names were individual constants. But it can easily be modified to accommodate the view ([4]) that they are predicates.

[2] '...' is eliminable. See Peacocke [12] n. 2 to see how.

[3] This is a different suggestion from the one which Baldwin discusses in [2]. He argues against giving explicitly metalinguistic construals of relational saying sentences, e.g. 'Ralph believes of $<$Tom,Dick$>$ that it satisfies "x_1 spies on x_2"'. (I suspect he attacks a straw man.)

seen to be disposed to speak of them. (There may prove to be other forms
of speech which force us to credit such a disposition to speakers, and, if so,
that would affect the view of the present case.) Perhaps it will be claimed
that it is quite indeterminate what kinds of thing the *n*-tuples introduced
above could be identified with, so that they may as well be identified with
sequences. However, the force of calling the *n*-tuples *primitive* was
precisely to suggest that we need not identify them with any theoretical
entities. In any case, it cannot be right to deal with a problem of under-
determination by settling for an account which goes clean beyond all
the evidence.

It is for these reasons that I would resist the accounts of Loar
([10] p. 57–8) and Temin ([16] p. 305), and that I doubt that Wallace in
[17] gave the logical forms of belief sentences. (He never said he did.)

Wallace represents notional belief sentences as expressing relations
between persons, times, propositions and *the null sequence*. One must
surely show some connection between the notional and relational
'believe', and Wallace's proposal affords a means of seeing one as a
special case of the other. But, if logical form is our concern, his repre-
sentations of the notional find our intuitive discomfort at the intro-
duction of sequences at its strongest. On my account there is no question
of the words 'believe' or 'say' each having multiple senses, though a dif-
ference between the relational and notional is marked none the less: the
'of' relation doesn't feature in the notional case. But then in the notional
case there is nothing the utterance or state reported was said to be of.

What led Wallace to his view of the notional—the suggestion that
belief parallels *satisfaction*—can perhaps be seen as a suggestion about one
way in which theories of belief and of speech are linked. It seems evident
that a theory of belief will mirror a theory of speech in crucial respects,
for the part of a theory of speech which is semantics must yield under-
standing of propositional acts and states of mind alike. We might then
expect to find parallels, as I have here, between sentences which con-
tain 'believe' and sentences which contain 'say': it is those sentences at
least that our theories of belief and speech must illuminate. But it is not
clear that language users who report belief and speech need have know-
ledge of anything more than the palest shadows of our philosophical
theories about them.[1]

Newnham College, Cambridge   © JENNIFER HORNSBY 1977

  [1] Something which Colin McGinn once said to me caused the thought which caused me
to write this paper.

REFERENCES
[1] Arnaud, Richard B. 'Sentence, Utterance and Samesayer', *Nous* X (1976) 283–304.
[2] Baldwin, Thomas. 'Quantification, Modality and Indirect Speech' in *Meaning Reference
and Necessity*, ed. Simon Blackburn, (Cambridge University Press: 1975).

[3] Burge, Tyler. 'Demonstrative Constructions, Reference and Truth', *Journal of Philosophy* 71 (1974) 205–223.

[4] — — — 'Reference and Proper Names', *Journal of Philosophy* 70 (1973) 425–439.

[5] Davidson, Donald. 'On Saying That' in *Words and Objections, Essays on the Work of W. V. Quine*, eds. D. Davidson and J. Hintikka (Dordrecht: Reidel, 1969).

[6] — — — 'The Logical Form of Action Sentences' in *The Logic of Decision and Action*, ed. N. Rescher (University of Pittsburgh Press: 1966).

[7] Davidson, D. and Harman, G., editors, 'Introduction' to *The Logic of Grammar*, (Dickenson: 1975).

[8] Hornsby, Jennifer. 'Singular Terms in Contexts of Propositional Attitude', *Mind LXXXV* (1977).

[9] Kaplan, David. 'Quantifying In', *Synthese* 19 (1968–9) 178–214.

[10] Loar, Brian. 'Reference and Propositional Attitudes', *Philosophical Review LXXXI* (1972) 43–62.

[11] McFeteridge, Ian. 'Propositions and Davidson's account of Indirect Discourse', *Proceedings of the Aristotelian Society LXXVI* (1975–6) 131–145.

[12] Peacocke, Christopher. 'An Appendix to David Wiggins' "Note"' in *Truth and Meaning*, eds. Gareth Evans and John McDowell (Clarendon Press: Oxford 1976).

[13] Quine, W. V. 'Quantifiers and Propositional Attitudes' *Journal of Philosophy* 53 (1956) 177–87.

[14] — — — 'Reply' to Davidson [5], *op. cit.*

[15] Schiffer, Stephen. 'A Paradox of Desire', *American Philosophical Quarterly* 13 (1976) 195–203.

[16] Temin, Marc. 'The Relational Sense of Indirect Discourse'. *Journal of Philosophy* 72 (1975) 287–306.

[17] Wallace, John. 'Belief and Satisfaction', *Nous VI* (1972) 85–95.

ANIMAL RIGHTS

By R. G. FREY

I

THE question of whether animals possess rights[1] is once again topical, largely as a result of the recent surge of interest in animal welfare and in the moral *pros* and *cons* of eating animals and using them in scientific research.[2] If animals do have rights, then the case for eating and experimenting upon them, especially when other alternatives are available, is going to have to be that much stronger; and those who engage in and support these practices are going to be increasingly beleaguered. Animal rights may not give vegetarians and animal liberationists all that they want, but the existence of such rights would unquestionably strengthen the cases of both camps. Arguments to show that animals do have rights, therefore, are at a premium. In what follows, I want to suggest that the most important such argument fails.

II

The argument in question consists in using the cases of babies and the severely mentally-enfeebled to force the inclusion of animals within the class of right-holders. Those who use the argument proceed this way: they first cite the many and various criteria by which philosophers and others have tried to show why human beings possess rights but animals do not, and then claim of each and every one of these criteria that it would exclude babies and the severely mentally-enfeebled as right-holders; since we all agree that babies and the severely mentally-enfeebled do have rights, each and every one of the criteria must be rejected as a criterion for the possession of rights. The form of the argument, then, is as follows:

> (1) Each and every criterion for the possession of rights that excludes animals from the class of right-holders also excludes babies and the severely mentally-enfeebled from the class of right-holders;

> (2) Babies and the severely mentally-enfeebled, however, do have rights and so fall within the class of right-holders;

[1] I refer here and throughout this paper, and in this I align myself with all the current writing on the subject of animal rights, to moral rights.

[2] See, for example, J. Feinberg, 'The Rights of Animals and Future Generations', in W. Blackstone (ed.), *Philosophy and Environmental Crisis* (Athens, Georgia, 1974); S. & R. Godlovitch, J. Harris (eds.), *Animals, Men and Morals* (London, 1971); A. Linzey, *Animal Rights* (London, 1976); T. Regan, 'The Moral Basis of Vegetarianism', *Canadian Journal of Philosophy*, vol. V, 1975, pp. 181–214; T. Regan, P. Singer (eds.), *Animal Rights and Human Obligations* (Englewood Cliffs, New Jersey, 1976); P. Singer, 'Animal Liberation', *The New York Review of Books*, vol. XX, no. 5, April 5, 1973, pp, 17–21; P. Singer, *Animal Liberation* (London, 1976).

(3) Therefore, each and every one of these animal-excluding criteria must be rejected as a criterion for the possession of rights.

Obviously, this argument is essentially negative and indirect, in that it does not aim so much to establish directly the positive thesis that animals have rights as to establish the negative thesis that animal-excluding criteria for the possession of rights will not do, since they exclude as well babies and the severely mentally-enfeebled. (Of course, the implication of the negative thesis is that, if we go on to adopt a criterion for the possession of rights that *includes* babies and the severely mentally-enfeebled within the class of right-holders, then it will include animals within the class of right-holders.) Thus, for example, rationality as a criterion must be discarded, for otherwise we obtain a singularly objectionable result:

> If we accord moral rights on the basis of rationality, what of the status of newly born children, "low grade" mental patients, "intellectual cabbages" and so on? Logically, accepting this criterion, they must have no, or diminished, moral rights.[1]

Instances of this argument abound, and by means of it possession of rationality or of a language, the recognition and discharge of moral obligations, the possession of a culture, the acceptance of and participation within societal and communal relationships, the possession of interests (where this connotes that something S is in one's interest, that one cares or exhibits concern about S, and perhaps that one is prepared to do something about and even to think that one ought to do something about S), etc., are all rejected as criteria for the possession of rights.

I want to suggest that the present argument does not work. It hinges upon premiss (2), that is, upon treating the cases of babies and the severely mentally-enfeebled as open and shut so far as the possession of rights is concerned. Premiss (2), however, is not obvious and requires defence; but the best defences of it, *if they stand at all*, specifically exclude animals from the class of right-holders. Therefore, either premiss (2) cannot be defended or else premiss (1)—that every animal-excluding criterion for the possession of rights also excludes babies and the severely mentally-enfeebled—is false; either way, this important argument for animal rights fails.

III

Is it so very clear that babies and the severely mentally-enfeebled do have rights—do have rights, that is, without the addition of *further arguments* which themselves exclude animals as right-holders?

For example, consider again the rationality requirement:

[1] A. Linzey, *Animal Rights* (London, 1976), p. 24.

Only beings which are rational possess rights.

Given a suitably restrictive analysis of rationality, babies and the severely mentally-enfeebled will be excluded from the class of rational beings and so from the class of right-holders; on this requirement, they simply have no rights. Since, upon the same analysis of rationality, animals also are not rational, it follows that they have no rights either.

Now there are three arguments by which one might try to include babies and the severely mentally-enfeebled within the class of right-holders and so to defend premiss (2); each in turn, however, specifically excludes Fido from this class.

(1) One might try to include the baby by means of the *potentiality argument*: the baby is potentially rational. Of course, the baby is not at present rational, and if actual rationality is insisted upon, then the baby has no rights. On the other hand, the potentiality argument does separate the case of the baby from that of Fido, who is not conceded even potential rationality.

(2) One might try to include the severely mentally-enfeebled by means of the *similarity argument*: in all other respects except rationality and perhaps certain mental accomplishments, the severely mentally-enfeebled betray strong similarities to other members of our species, and it would and does offend our species horribly to deprive such similar creatures of rights. If this argument is rejected, on the ground that rationality is the requirement for possessing rights and other similarities are beside the point, then the severely mentally-enfeebled do not have rights. On the other hand, the similarity argument does separate the severely mentally-enfeebled from Fido, who does not bear anything like (even) sufficient physical similarities to ourselves to warrant similar inclusion.

Animal liberationists, of course, will also object to the similarity argument on the ground that it smacks strongly of speciesism. For it does enshrine, if not advocate, active discrimination against other species in favour of our own.

(3) One might try to include both babies and the severely mentally-enfeebled by means of the *religious argument*: babies and the severely mentally-enfeebled possess immortal souls. If this argument is rejected, on the ground that, even if they possess immortal souls, beings must also possess rationality in order to have rights, or on the ground that there is no good evidence for the existence of such souls, then neither babies nor the severely mentally-enfeebled possess rights. On the other hand, the religious argument does separate both from Fido, who is not conceded an immortal soul by the argument's proponents.

The upshot is this. Unless one of these three arguments is accepted, we have no basis upon which to differentiate the cases of babies and the

severely mentally-enfeebled from that of animals; and if all three of the arguments *are rejected*, and there are serious objections to each, then it follows on the requirement under consideration that babies, the severely mentally-enfeebled and animals are alike in not being right-holders. In other words, the best defences of premiss (2) collapse, with the result that the premiss cannot sustain the weight put upon it. If, however, one of these three arguments *is accepted*, so that babies and the severely mentally-enfeebled are held to fall within the class of right-holders, then we find that that argument itself specifically *excludes* animals from the class of right-holders. In other words, premiss (1) is false, since not every animal-excluding criterion for the possession of rights excludes babies and the severely mentally-enfeebled.

For these reasons, then, I conclude that either premiss (2) cannot be defended or else premiss (1) is simply false, so that, in either case, this most important argument in behalf of animal rights fails.[1]

University of Liverpool   © R. G. FREY 1977

CAN THINGS OF DIFFERENT NATURAL KINDS BE EXACTLY ALIKE?

By JOHN TIENSON

IS it possible for two things to be exactly alike and for one and only one of them to be a cow? One's first response to this question, I suppose, is likely to be that it is not possible. After all, if scientists were to create an animal genetically identical to Bossie in the laboratory, we would surely say they had created a cow. Nevertheless, I believe our intuitions, properly nudged, lead to a positive answer to this question. A similar point concerning artificial kinds is certainly correct. A door and a table top, for example, could in principle be identical in physical structure.

If it is possible for things that are exactly alike to be of different natural kinds, it follows that being of a certain kind cannot be identified with having a certain physical or genetic structure. Putnam has shown[1] that the mental states of a speaker together with the physical nature of a thing are not sufficient to determine whether a natural kind term used by that speaker applies truly to the thing. If I am right, in some cases at least, the nature of the thing together with the nature of things to which the term has been applied in the past are still not sufficient to determine whether the term truly applies. We are faced, then, with the question of what does determine whether such a term applies to a thing. Section V goes into these points in a bit more detail, and in Section VI I offer a brief suggestion about the kind of answer I think we should give to this question. In Section I, I attempt to clarify the title question enough to get on with the story. Section II introduces an example which, hopefully, will lead us to say that an animal that is exactly like a certain cow is not a cow. In Section III, I will argue that anything else one might want to say about the example is unsatisfactory. Section IV considers briefly the range of terms for which such examples can be constructed.

I

Obviously, if x and y share all relational properties, they are the same thing, since they share the property of being identical with x. Further, if physical things like cows share all spatial and temporal relations, they are once again the same thing. Let us not count spatio-temporal location among *monadic* properties. And let us agree to count all aspects of the

internal structure of a thing among its monadic properties. Finally, let us include among the monadic properties complex "properties" that are definable solely in terms of monadic properties. Thus, if "bachelor" is definable as unmarried male, and being male and being unmarried are regarded as monadic properties, then being a bachelor will be as well. Then we can ask: Can two things have exactly the same monadic properties and be different kinds of things?

Of course, if being a cow is counted among the monadic properties, the answer to this question is again trivially negative. One of the upshots of this discussion is that being a cow probably should not be considered a property of an individual in the way that its specific size, shape, internal structure and so forth are, nor should it be thought of as definable in terms of such properties. But we can remain neutral on this and ask,

> (1) Can two things share all of their monadic properties other than that of being a cow and those definitive of being a cow, and one and only one of them be a cow?

Finally, complex organisms change. It is too much to expect that two things like cows should literally be exactly alike throughout much of their careers. But we can imagine them to be at least as much alike as identical twins, genetically identical, indistinguishable in appearance, and completely alike in physical structure at least once in their lives. Let them, that is, be as much alike as two organisms can conceivably be.

II

Bossie is a cow, on earth, in America. Her parents were cows; she has given birth to calves. Anyone seeing her in the barn would recognize her as a cow. Bessie, on the other hand, is a Martian guelph. Her parents were guelphs, and she has offspring that are guelphs. Any Martian seeing Bessie in the field would recognize her as a guelph. Bossie and Bessie are exactly alike. If either were secretly replaced by the other, an examination would not reveal the exchange. Had either replaced the other at the moment of conception of her offpsring, the resulting offspring could have been indistinguishable from those actually conceived. (I will not speculate on the species of these possible offspring.)

But most cows are quite unlike any guelphs, and on Mars would not be taken for guelphs. Likewise, most guelphs would not be taken, or mistaken, for cows. There is a considerable range of variation among both cows and guelphs, in size, markings, temperament, etc., and of course, also genetically. But there is a slight overlap of these ranges of variation, of which Bossie and Bessie constitute a striking instance. Bessie is within the range of variation of normal guelphs—she is no

freak. Similarly for Bossie. Indeed, and this is important, Bossie stands in just the same relation to the range of variation of normal cows as Bessie stands to the range of variation of normal guelphs. We may also suppose that interbreeding between cows and guelphs is in general not possible, and that it is possible only for individuals falling within or near the overlap of ranges of individual variation.

What we imagine, then, are two completely distinct biological communities, with completely distinct evolutionary histories. Within each of these communities we imagine what we would certainly regard as a species. And we suppose that some members of one of these separately evolved species are very similar to or even exactly like some members of the other, although most members of either of the species are not very much like members of the other. It might be useful here to think of the species of dogs (*canis familiaris*) with its diverse breeds. Imagine, then, that evolution on another planet has produced a species that is just as diverse, and some mongrel runts of which are just like Saint Bernards.

III

It seems clear to me that we should say that Bossie and Bessie are not members of the same species, that Bossie is a cow and Bessie is a guelph, and that cows are not guelphs. But in case there is a temptation to say that they are members of the same species, let us consider what we would have to say about the rest of the cows and guelphs. There are five possibilities, none satisfactory.

First and second, Bossie and Bessie are both cows and not guelphs, or vice versa. But as we have constructed the example, it is symmetrical in all relevant respects. There is no reason for Bessie's being a cow and not a guelph (or for Bossie's being a guelph and not a cow). This is not an epistemological point, that we could not tell which she is. It is a question of what makes her a cow. It cannot be her similarity to cows that makes Bessie (the Martian) a cow, because she has the same relations of similarity to guelphs, and so would have to be a guelph as well (the third possibility). But if it is something other than her similarity to cows that makes her a cow, we would have to say that she would have been a cow (and not a guelph) even if evolution on earth had taken a completely different course and there had never been any cows, or even mammals, on earth. This, clearly, we do not want to say.

Third possibility, cows and guelphs are one species, with different names in different places. There are two reasons for not taking this course. First, most cows cannot interbreed with most guelphs. In typical cases of one species in different areas, interbreeding is universally or almost universally possible, and the possibility of interbreeding is one of the primary criteria for sameness of species. Second, cows and guelphs have completely distinct evolutionary histories. Were their

ranges of individual variations nearly identical, there might be some reason to speak of separate evolution of the same species, but that is not the case here. There is even less reason to say they constitute a single species than there would be to say that dogs and Tasmanian wolves (which are marsupials) constitute a single species if it happened that Tasmanian wolves were nor merely dog-like in appearance, but occasionally indistinguishable from dogs.

Fourth possibility, cows and guelphs are two different species, and Bossie and Bessie belong to both of them. One trouble with this line is that it seems to be a fundamental aspect of the logic of natural kind terms that they are mutually exclusive at a given level of taxonomy. 'It is a lion; therefore, it is not a tiger', is an acceptable inference. 'Bossie is a cow; therefore, she is not a guelph', should be equally acceptable. More important, we would lose transitivity of sameness of species in a most unwelcome manner. Bessie is certainly of the same species as her parents, and Bossie belongs to the same species as her parents. On this line, Bossie and Bessie would be members of the same species, but Bossie's parents would not be members of the same species as Bessie's parents. Worse, Bessie's parents and offspring are members of one species, guelph. It seems clear that Bessie is not a member of some species of which her parents and offspring are not members.

Fifth possibility, there are three or more species involved, one of which comprises the overlap of ranges of variation of "cows" and "guelphs". But again, we can construct the example so that Bossie's parents and offspring are not in this overlap or even very near it. Then, on this alternative Bossie would not be a member of the same species as her parents and offspring. But it seems clear that she is.

IV

Thus, it appears that things can be exactly alike in physical structure and be different kinds of things in the sense of falling under different natural kind terms, at least when those terms are terms for biological species. The door and table top mentioned in the first paragraph show that things can be exactly alike and be different kinds of artifacts. It is possible to imagine similar, though more laboured, examples for conventional activities. That is, people or groups of people could go through exactly the same physical and mental processes and yet, in different cultural settings, be engaged in different activities.

There are, however, natural kind terms that appear to resist such examples, in particular, terms for chemical and physical kinds. If a lump of stuff on another planet were structurally identical with an ice cube in my refrigerator, it would be ice. What makes 'ice' different from 'cow' in this respect is not that it is a term of a kind of stuff. If the milk of a certain breed of guelph became indistinguishable from peanut butter

after a few days' exposure to air, it would still not be peanut butter. On the other hand, anything exactly like a proton is a proton, and anything exactly like a helium atom is a helium atom, no matter how it came into existence. Apparently, for relatively basic kinds only, possession of certain monadic properties, including internal structure, is sufficient to determine natural kind.

V

In formulating (1) we left open the question of whether being a cow was to be regarded as a property of a thing. Any view which posits something objective corresponding to natural kind terms, which all and only members of a kind have, and by virtue of which they are of that kind, can reasonably be called realism with respect to natural kind terms.[1] We can, with some historical justification, call such posited objective determiners of natural kinds essences. Most of us today, I suppose, are inclined to think of natural kinds as determined by some aspect of the structure of members or bits of the kind, molecular and atomic structure, for example, in the case of chemical and physical kinds. For biological species, physiological or (more likely) genetic structure is the best candidate for the essence of the species. But if it is possible for things of different kinds to be identical in monadic properties including internal structure, as I have argued, essences as determiners of natural kinds cannot be identified with such structures. There is no class or kind of structure which is necessarily shared by all and (in particular) only members of one kind.

It does not follow that there are no such things as essences that determine species membership. But there are formidable tasks for a view which posits essences that are distinct from structures. It must explain the relationship of the essence to the structure of individual members of species (since we could have different essences and identical structure). And it would have to explain how such essences are realized in individuals. It is also difficult to see how such entities could play a role in an explanation of our use of natural kind terms.

If there is nothing objective in an individual that makes it a member of a biological kind, what does make Bossie a cow? We cannot simply say that she is a cow because she stands in certain biological relations to some *cows*. For there must be something which makes *them* cows. But there was nothing very special about Bossie among cows. If there is a problem about what makes Bossie a cow, there is just as much of a problem about what makes each of her relatives a cow.

---

[1] I take it for granted that there must be some bojective cases of sameness of kind (cases, that is, that do not depend on our capacity to attribute sameness), and hence some objective determiners of sameness of kind or likeness, i.e. some universals. The question here is whether there are any universals corresponding to natural kind terms.

Still, Bossie's biological relations with other cows are important in making them animals of the same kind. As Putnam points out (MM, 141f; M&R, 702), when we give an ostensive definition of a natural kind term like 'water' we presuppose 'that the body of liquid I am pointing to bears a certain sameness relation (say, . . . x is the $same_L$ as y) to most of the stuff I and other speakers in my linguistic community have on other occasions called "water" '. *Given* that this presupposition is fulfilled, 'the necessary and sufficient condition for being water is bearing the relation $same_L$ to the stuff in the glass'. Now, in the case of water, and chemical and physical kinds in general, the relevant relation of sameness of kind is determined by certain kinds of monadic properties. Putnam's arguments show that the relevant monadic properties are not superficially observable ones, but hidden, internal ones which determine the observable properties. As Putnam rightly emphasizes, we need not know what these properties might be like. Speakers of English before the advent of chemical theory were in a position to know that it was something about the internal nature (epistemically, it might have turned out not to be structure) of the stuff that made it water.

If this is correct, then (as Putnam argues at length in MM and M&R) the mental states of speakers and the nature of a thing are not in general sufficient to determine whether a particular natural kind term applies to the thing. The speaker presupposes a relation of sameness of kind. But a particular speaker may not know, and at a given time no speaker may know, what this relation consists in. In such a case, the speaker's present mental state will not determine what a thing must be like to stand in the relevant relation of sameness of kind to things to which the term has been applied in the past, and hence will not determine what the thing must be like to be of that kind.

Putnam assumes that hidden, internal properties determine sameness of kind for all natural kinds. But our example shows that there are natural kind terms that do not work this way. Here, sameness of monadic properties, observable or hidden, does not imply sameness of kind. There is a sense in which Bossie is the same kind of thing as her parents (a cow), but is not the same kind of thing as the qualitatively identical Bessie. Apparently, there are at least two classes of natural kind terms. For one class, we regard certain kinds of monadic properties as sufficient for determining the relation of sameness of kind that is relevant for their application. For the other class of natural kind terms, we regard the origin of a thing or batch of stuff as an important factor in determining the relevant relation of sameness of kind. For things to which we apply such terms, we will generally not count things as being of the same kind unless we can also regard them as coming from the same kind of source.

VI

When one learns a natural kind term, he presumes that a kind has been identified and that term applied to it. He need not encounter a sample of the kind himself, but in typical cases others in his linguistic community will have, and there will be a historical link between their use of the term for certain things and his present acquisition of the term. This presumption that a kind has been identified requires that there be a relation of sameness of kind that holds among (most of) the samples encountered in the past and others which may be encountered in the future. Now, there are in the world individuals and batches of stuff, each with its internal constitution, and there are similarities between individuals or their constitutions. And these individuals and batches of stuff have come about in various ways. There are no relations of sameness of kind over and above these similarities and causal relations. Sameness of natural kind is, indeed, a matter of what we *regard* as making things the same kind, of how we use the similarities and causal relations to classify things. Of course, the world must co-operate. The things we classify must have enough interesting similarities in properties and origins for us to be able to regard them as falling into kinds. What emerges from our example is that there are two discernible ways in which we regard the facts as determining sameness of kind, and furthermore, that we can know this—that there are two ways—before we know what the facts are.

Why should we classify things in two different ways? There is a significant difference between the two cases in the degree of similarity of samples of the kind. Members of the same species are not exactly alike, genetically or in any other way. Atoms and molecules of the same kind, on the other hand, are so nearly identical as to be indistinguishable. And there is a single specification of their structure that is applicable to all and only members of a single kind, which is not in general the case for biological kinds. One might even see in this sufficient reason for saying that chemical and physical kinds exist independent of our classifying—as certain kinds of structures—while biological kinds do not.[1] I do not think, however, that this difference in degree of similarity can be the reason we have two modes of classifying natural kinds. The similarities that are crucial in determining sameness of kind are, after all, hidden similarities. But, I think there can be no doubt that these two ways of classifying existed before it was known how the degrees of similarity would turn out. If there is anything in human experience that contributes to our making this distinction, I should think it would be the fact that the origin of members of biological kinds

---

[1] For a formulation of a logic of natural kinds in which the cases where some, some kinds of, all, or none of our natural kind terms correspond to real kinds are easily handled, cf. Nino Cocchiarella, 'On the Logic of Natural Kinds', *Journal for the Philosophy of Science* (forthcoming).

(and the substances derived from them) is so much better known and so much more important to us.

At any rate, according to our intuitions concerning biological kinds, it is possible for things which are exactly alike to be regarded as belonging to different kinds. For this to be the case, internal constitution cannot be the sole determiner of kind membership.

California State University, Northridge   © JOHN TIENSON 1977

MUST WE ACCEPT EITHER THE CONSERVATIVE OR THE LIBERAL VIEW ON ABORTION?

By HUGH V. McLACHLAN

I

WHEN considering the morality of abortion, it has been common to focus attention upon the ontological status of the fetus and, specifically, to ask whether or not the fetus is a person. Indeed, as Hare points out, the question of:

> . . . whether the fetus is a person, has been so universally popular that in many of the writings it is assumed that this question is the key to the whole problem. The reason for this is easy to see; if there is a well-established moral principle that the intentional killing of other innocent persons is always murder, and therefore wrong, it looks as if an easy way to determine whether it is wrong to kill fetuses is to determine whether they are persons, and thus settle once for all whether they are subsumable under the principle (R. M. Hare, 'Abortion and the Golden Rule', *Philosophy and Public Affairs*, vol. 4, 1975, p. 204).

If we approach the abortion issue in this way, then it appears that abortion is never, or almost never, justifiable or else never in need of justification: 'For, if a human fetus is a person, one is inclined to say that, in general, one would be justified in killing it only to save the life of the mother. Such is the extreme conservative position. On the other hand, if the fetus is not a person, how can it be seriously wrong to destroy it?' (M. Tooley, 'Abortion and Infanticide', *Philosophy and Public Affairs*, vol. 2, 1972, pp. 38–39). That it is never wrong to destroy it is the extreme liberal position. There appears to be no plausible middle ground between the extreme conservative and the extreme liberal views on abortion. I want to argue that this appearance is deceptive and that the deception is produced by a misrepresentation of the ontological status of the fetus and of the importance of this status.

What is the ontological status of the fetus? According to Hare, this is a moral question. For reasons that need not concern us here, he thinks that ' . . . to say that the fetus becomes a person at conception, or at quickening, or at birth, or whenever takes your fancy . . . is inescapably a moral decision for which we have to have moral reasons' (Hare, p. 205). An opposing point of view is presented by, for example, Werner. He distinguishes a 'human being' by which term he means ' . . . a member of the biological species *homo sapiens*' from a 'person' which is said to be ' . . . a fully fledged member of a human community, someone having a developed concept of *self*, memories, a language and/or moral obligations as well as moral rights . . .' (R. Werner, 'Hare on Abortion', ANALYSIS 36.4, 1976, p. 178). He concludes that fetuses are not persons but are human beings: they are ' . . . the live, growing uterine offspring of human beings and are not members of any other plant or animal species' (Werner, p. 179). Both writers have committed a simple category error which commonly occurs in the abortion debate. Is the fetus a person or is it a human being? When does the fetus become a person or a human being? A fetus is neither a person nor a human being; a fetus never becomes a person or a human being: it is or might become the body of one. In at least one respect, the question of the ontological status of the fetus is quite unproblematical and non-moral. A fetus is a partially developed human body (I argue this in 'Moral Rights and Abortion', *Contemporary Review*, vol. 228, no. 1325, June 1976, pp. 323–8).

People sometimes ask what, if any, are the rights of the fetus and this is a curious question. We never attribute rights to fully developed bodies so why should we attribute them to partially developed ones? After all, we would not say that my arms or legs or any other part of my body, or, indeed, my entire body had rights. It is, rather, me who has rights. Rights are held by people, not by their bodies. Should we therefore conclude that the extreme liberal position is firmly established and that since a fetus cannot have rights, abortions are not in need of justification when the prospective mothers desire them? We can validly reach this conclusion only by making two assumptions which can be challenged. Consider Tooley's claim that: 'Settling the issue of the morality of abortion and infanticide will involve answering the following questions: What properties must something have to be a person, i.e. to have a serious right to life? At what point in the development of a member of the species Homo sapiens does the organism possess the properties that make it a person?' (Tooley, p. 43). The assumptions to which I refer are being made here by Tooley. Firstly, it is assumed not only that rights are due only to persons but that such persons must be living, actual persons. Secondly, it is assumed that in order for us to have duties regarding abortion, there must be someone possessing corresponding rights.

Let us deal with these assumptions in turn. Suppose we promise a friend that we shall never reveal a particular secret. Are we released from our duty not to break our trust on our friend's death? Can we not say that he still has a right not to have his confidence betrayed and that we have a duty not to betray it? Suppose we promise to place flowers on his grave. It seems to me that our dead friend has a right to have flowers placed on his grave and that we have a corresponding duty to place them there. And such duties to the dead need not arise through contract or personal relationships. I never met Wittgenstein. Yet, I would suggest that I have duties towards the dead man. For example, I have a duty not to slander him and it is not an obvious abuse of language to say that he has a right not to be slandered. To say the least, it is not self-evident that only actual living persons have rights. We might say that the person whose body a fetus might develop into has rights and that, correspondingly, we have duties towards him whether or not he is an actual living person or ever will become one. We can have duties towards the person who might or will become the occupier of a particular body if a fetus develops into a fully formed human body. If I am correct in maintaining that a non-existent person can have rights and that we can have duties towards him, then the question of the ontological status of the fetus need not be considered to be critically important and certainly not the central issue in the abortion debate. We need not be particularly concerned with the question whether the fetus is the partially formed body body of an actual person or of a human being or the question at what precise stage a fetus, or, indeed, the body of a baby becomes that of an actual person or human being.

We might even talk of our having duties towards persons whose bodies are not yet fetuses. Suppose someone knew that if he were to have sex with his wife at a particular time then the human being who would be produced would live in abject misery because of, say, some physical ailment. He would surely have a duty not to have sex at that particular time and why should we shy from saying that the miserable person who would be produced has a right against him that he does not have sex at that particular time? If talk of rights with non-existent holders seems peculiar then it could be said simply that he has a duty not to have sex at that particular time and that no one holds a corresponding right. This leads us to the second assumption to be considered.

It is doubtful whether on every occasion when we have a duty to do X, in the sense that we ought to do X, that X be done is some particular person's right. Suppose that we were living in a country where many people around us were starving and that we had far more food than we needed for both our present and our future purposes. We would surely have a duty to give some of our food away, in the sense that we ought to do so, even if it were the case that no particular person had a right to

receive food from us. We might have enough spare food to feed only one person and have a duty to feed one starving person where no starving person had any more right to receive the food than had any other. Consider our conservation policies or our moral concern at their relative absence. Suppose that all the world's natural resources will run out in one hundred years if we continue with our present rate of consumption and that the earth, and all other planets, will be unfit for human habitation. Should we be morally concerned at this possible state? Would we not think it immoral for us avoidably to act in such a way that future human life would be precluded? Even if we were to deny that the members of potential generations have rights, I am sure that most of us would assert that we have a duty not to dissipate natural resources just as we would have a duty not to waste food in the company of starving people even if no one is the holder of a corresponding right. Once again, if we can have moral duties without there being holders of corresponding rights, then the question of the ontological status of the fetus and of whether this status is such that the fetus could reasonably be considered to be the holder of rights need not be considered central or critical in determining the morality of abortion.

II

Before considering an alternative to the extreme liberal and extreme conservative views on abortion, I want to discuss some further arguments for the liberal view.

It might be thought that if the fetus is simply a partially formed human body then it is the property of the woman who carries it and that she is morally entitled to do what she wants with it. It might be thought that abortion is no more a moral issue than tooth extraction. Is it not the case, as Thomson claims, that '. . . the mother and the unborn child are not like two tenants in a small house which has, by an unfortunate mistake, been rented to both: the mother *owns* the house'? (J. J. Thomson, 'A Defense of Abortion', *Philosophy and Public Affairs*, vol. 2, 1972, p. 53). It is curious to talk of ownership in relation to any bodies, most of all fetuses. For instance, when we own a house we can sell it and cease to own it, but, even if we were prostitutes or slaves, our bodies would, indeed must, in an obvious sense remain our own. Is Thomson talking about legal ownership or what we might call 'logical ownership'; or is some notion of 'moral ownership' involved? If the fetus is that of the mother, is it not also that of the father? Furthermore, could the fetus not be thought to 'belong' to the person into whose body it might develop? It is his body after all. It is not clear who, if anyone, is the sole 'owner' of the fetus and it is not clear in what sense, if any, it is appropriate to talk of ownership in relation to the fetus.

Nevertheless, it might be thought that the fetus is part of a woman's

body. What would follow from this? Is a woman entitled to do with it as she pleases because of her '. . . right to decide what happens in and to her body which everyone seems ready to grant'? (Thomson, p. 50). It is not clear that anyone has such a right. A right to decide what happens to one's body might be the absence of duties towards other people to treat our bodies in particular ways. But we are surely not free of such duties. Would a husband not be failing in his duty towards his wife if he took a drug such as heroin to such an extent the he was never capable of sex and failing in his duty towards his children if he abused his body in such a way that it was never capable of work? The fact that a fetus is part of a woman's body does not preclude her having duties regarding its disposal towards the person whose body the fetus might become. To have a right to decide what happens to one's body might mean that there is nothing which one ought to do or ought not to do with one's body; it might mean that there is an absence of all duties regarding its use and not just of duties with corresponding rights. Even if no one had a right to our bodies or a right to a particular performance from them, we might still have duties regarding our bodies. We might have a duty to keep our bodies healthy even if no one has a right that our bodies are healthy. A woman might 'own' a fetus totally in the sense that no one has a right that it is treated in any particular way and yet have duties regarding its disposal just as we would have duties regarding the disposal of our own food in a land of starving people. For example, we might imagine a situation where the birth-rate had declined to such an extent that the human race was in serious danger of extinction. A woman then might have a duty to have a baby rather than an abortion even though no one else had any rights regarding the treatment of her fetus. We might think that it is morally good to perpetuate human life even if we did not think that any particular person had a right to be born. In some other situation, a woman might be morally obliged to have an abortion even if it were no one's right that she had one.

The following objection is raised by Tooley against the view that currently non-existent persons can have rights. He asks us to imagine a situation where:

> . . . a chemical [is] discovered which when injected into the brain of a kitten would cause the kitten to develop into a cat possessing a brain of the sort possessed by humans and consequently into a cat having all the psychological capabilities characteristic of adult humans. Such cats would be able to think, to use language and so on. Now it would surely be morally indefensible in such a situation to ascribe a serious right to life to members of the species Homo sapiens without also ascribing it to cats that have undergone such a process of development; there would be no morally significant differences (Tooley, pp. 60-1).

He then assumes that if a kitten had been injected with a such a chemical it could not be morally wrong to prevent the kitten's developing into a

cat with the mentioned human-like qualities and concludes: 'But if it is not seriously wrong to destroy an injected kitten which will naturally develop the properties that bestow a right to life, neither can it be seriously wrong to destroy a member of Homo sapiens which lacks such properties, but will naturally come to have them' (Tooley, p. 61). According to Tooley, abortions are not in need of justification and the extreme liberal position is established.

One might reject Tooley's conclusion and insist that it would require justification to kill such injected kittens. Tooley argues that 'It is hard to believe that there is much to be said for the . . . moral claim . . . that in a world in which kittens could be transformed into "rational animals" it would be seriously wrong to kill newborn kittens' (Tooley, p. 62). If he finds this moral claim implausible then this is surely a matter concerning his own psychology or moral judgment; the moral claim appears to me to be one which could rationally be held. Might it not, at least in some circumstances, be wrong wantonly to destroy kittens which would not develop into any thing other than normal cats?

Let us take Tooley's illustration to its logical conclusion. Suppose there were a chemical which when injected into the body of, say, a kitten would cause the body to become that of a person who was like other human persons in every respect other than genealogy. Suppose that a kitten had just been injected with the chemical. Tooley would claim that this kitten is on a moral par with a normal fetus. Well, we can concede this claim and say that the same moral consideration would be due to the person whose body that of the kitten would become as to any other person whose body developed in the more normal way. It might be that *no moral consideration* was yet due to any such persons and anyone who claimed that abortion or the killing of the imagined kitten are never in need of any moral justification would be committed to that view. Nevertheless, we are not logically bound to accept the view. We can rationally claim that moral respect and consideration are due to such persons even if we do not want to claim that in all or in any cases they have anything approaching a right to be born.

To the claim that non-existent people can have rights, it might be objected that since they cannot be identified, we cannot have duties towards them and consequently they cannot have rights. Hare considers such an objection and suggests that 'identify' is an ambiguous term. 'I can identify the next man to occupy my carrel at the library by describing him thus, but in another sense I cannot identify him because I have no idea who he is' (Hare, p. 220). He thinks that we can identify the supposed object of the duty to abort or not abort in the former although not in the latter sense: '. . . he is identifiable in the sense that reference can be made to him' (Hare, p. 220). For example, 'The person who will be born if these two people start their coitus in precisely five

minutes is identified by that description' (Hare, p. 220). Similarly, as I have done, we can talk of the person whose body a fetus might, will or would have become and that is to identify him. In any case, why should identification be thought important? Well, if I have a duty to give some-one a sum of money then it would seem that the person would have to be identifiable in the sense that he could be recognized as the holder of the right, picked out from a crowd as it were, since he would have to be (directly or indirectly) contacted by me in order for me to give him the money. If I could not contact him and know that I had contacted him and not someone else, then it could be argued that it would be impossible for me to fulfil my duty to give him the money and con-sequently that I do not have a duty to give him the money and he does not have a right against me to its receipt. But the situation is quite different in the case of our duties towards the unborn since it is possible for us to fulfil them without being able to recognize or contact the people involved. We might fulfil our obligations towards the members of future generations by failing to destroy the world without knowing who such people are.

III

We need adopt neither the extreme liberal nor the extreme conserva-tive position on abortion. We can hold that the fetus is neither a person nor a human being nor the body of an actual living person or living human being without being committed to the view that abortions are never in need of a justification. We can assume that it is not only actual living persons who have moral rights and/or that there can be moral duties which do not correspond to a person's rights. We might say that in some, although not necessarily all, cases there is either a duty to abort or a duty not to abort where this duty corresponds to a non-existent person's right and/or we might say that there is simply either a duty to abort or a duty not to abort. Although I have not refuted either the extreme liberal or the extreme conservative positions and have not sought to do so, I have tried to show some of the assumptions behind the positions and that they can be denied. Yet, an extreme liberal might still maintain that when a woman seeks an abortion, such abortion is not in need of a moral justification. He might say that although it is logically possible that non-existent persons have rights, they do not have them in fact; and that although it is logically possible that we have duties on other grounds either to promote or not to promote abortions, as it so happens we do not. He might say that even if non-existent persons have rights, their rights are so weak that we are always justified in not up-holding them if they conflict with a potential mother's wishes; and similarly that any other duties which there are pertaining to abortion are comparatively so trivial that a woman's wish to have an abortion will

always justify her having one. Similarly, an extreme conservative might maintain that, even though a fetus is not a person nor even the body of an actual living person, abortions are (almost) never justifiable either because of the strength of the rights of the non-existent persons involved or on some other grounds.

What position should we adopt in the abortion issue? I have not tried to answer that question and, apparently unlike some participants in the abortion debate, I consider this to be a straightforward moral question rather than a philosophical one. Between the various possible answers to this moral question, my comments are, I think, ethically neutral. Moral philosophy can indicate the range of possible answers to the question and what assumptions underlie them; moral judgment is required to decide what answers and assumptions are acceptable.

University of Glasgow   © HUGH V. McLACHLAN 1977

'WHATEVER ARISES FROM A JUST DISTRIBUTION BY JUST STEPS IS ITSELF JUST'

By EDWARD QUEST

AN important aspect of the entitlement theory of justice presented in Robert Nozick's *Anarchy, State, and Utopia* (New York, Basic Books, 1974) is the author's belief that the familiar non-fraudulent bilateral exchange is justice-preserving. How justice is preserved by such a transfer is illustrated supposedly by hypothesizing a fair distribution followed by a million petty 'capitalist acts between consenting adults' (each involves one quarter exchanged for the opportunity of watching Wilt Chamberlain play one game) such that the cumulative effect is a large transfer of money into the account of the star, Chamberlain. Each transfer was a just step, a step which could not transform a just distribution into a distribution which was not just, and so Nozick concludes, I think hastily, that the final distribution was just. His haste arises from his quick acceptance of the principle:

  N   'Whatever arises from a just distribution by just steps is itself just' (*op. cit.* p. 151).

I

N is not obvious, as may be seen by considering a version of Eubulides' "heap" paradox. Imagine a penniless beggar standing at a corner. A million people walk by within a year, each giving the beggar a

penny. If confronted with the questions, 'When exactly did this man cease to be poor?', 'Which penny was it which transformed him from being poor to no longer being poor?', one cannot provide answers free from arbitrariness. One's "ignorance" here stems from the inexactness of the concept of poverty, the absence of a precise borderline separating the poor from those who are not poor, not, of course, from any doubts as to whether or not the beggar was poverty-stricken at the end of the year. Once the inexactitude of the class of the poor is recognized, the intuition that a penny gift cannot be looked upon as having the capability of transforming a poor man into a man who is not poor will not be viewed as incompatible with the intuition that some series of penny gifts will not preserve poverty. Generally, if 'F' is inexact, it is not inconsistent to have a type of transfer which individually can never be pictured as changing an x which is F into an x which is not F even though a series of such transfers can be pictured as accomplishing this change. Thus, the intuition that a voluntary 25 cent exchange with Wilt Chamberlain could never be seen as transforming a just distribution into one which is not just (the intuition that such an exchange could not give rise to a legitimate complaint about the justice of the new distribution) is not sufficient for supporting the belief that any series of such exchanges will preserve justice: for if justice is inexact, a trivial exchange should not be expected to entitle anyone to make the non-trivial switch from 'the distribution is at present just' to 'now the distribution is not just'.

II

In presenting the Chamberlain example, Nozick challenges the reader to find some room for someone to complain that the distribution (D_2) brought about by the million trivial transfers is unjust.

> . . . Wilt Chamberlain winds up with $250,000, a much larger sum than the average income and larger even than anyone else has. Is he entitled to this income? Is this new distribution D_2 unjust? If so, why? There is *no* question about whether each of the people was entitled to the control over the resources they held in D_1; because that was the distribution (your favorite) that (for the purposes of argument) we assumed was acceptable. Each of these persons *chose* to give twenty-five cents of their money to Chamberlain. They could have spent it on going to the movies, or on candy bars, or on copies of *Dissent* magazine, or of *Monthly Review*. But they all, at least one million of them, converged on giving it to Wilt Chamberlain in exchange for watching him play basketball. If D_1 was a just distribution, and people voluntarily moved from it to D_2, transferring parts of their shares they were given under D_1 (what was it for if not to do something with?), isn't D_2 also just? If the people were entitled to dispose of the resources to which they were entitled (under D_1), didn't this include their being entitled to give it to, or exchange it with, Wilt Chamberlain? Can anyone else complain on grounds of justice? Each other person already has his legitimate share under D_1. Under D_1,

> there is nothing that anyone has that anyone else has a claim of justice against. After someone transfers something to Wilt Chamberlain, third parties *still* have their legitimate shares; *their* shares are not changed. By what process could such a transfer among two persons give rise to a legitimate claim of distributive justice on a portion of what was transferred, by a third party who has no claim of justice on any holding of the other *before* the transfer? (*ibid.*, pp. 161–2).

The suppressed argument in this passage concludes, I assume, with a variant of *N*.

> (a) If *X* chooses to transfer what *X* is entitled to transfer then *X* cannot legitimately complain about *X*'s transfer.
>
> (b) If *Y* is not deprived of anything *Y* is entitled to control *Y* cannot legitimately complain when another party *X* voluntarily transfers goods which *X* is entitled to transfer.

Consequently, any series of voluntary transfers stemming from a just distribution cannot give rise to a legitimate complaint.

The first premise is ambiguous. If the conditional is taken to be intuitive, the impossibility is seemingly so ephemeral that *X* can come to have a legitimate complaint at the end of a series of transfers of which his transfer is a member. If the impossibility of *X* having a legitimate complaint about his transfer is taken to be permanent, the conditional is counter-intuitive. Neither interpretation supports *N*.

Assuming *X* is not unfairly forced to choose from limited options, it is reasonable to insist that when *X* chooses to φ *X* is in no position to find φ-ing reprehensible. He has no reason to complain about what he is choosing to do. But this does not lead to that 'hobgoblin of little minds' —that *X* can have no reason to judge today that the action he chose yesterday was a mistake. Oedipus chose to have intercourse with Jocasta, he voluntarily slept with Jocasta; he then learned of his involuntary incest, and at this point he could find fault with what he had done. Generalizing from this understandable sequence, revaluations of voluntary actions can occur without any logical oddity as soon as one learns that what was chosen under one description can be redescribed correctly as an action which one would not have chosen voluntarily. Because such revaluations and the regret that accompanies them are based upon new information about the action, i.e. relevant features unknown when the choice occurred, they are quite distinct from the divided mind, the accompanying remorse, of moral weakness. In accounting for ordinary regret there is no reason to use the philosophically troublesome language involved in describing how a man intentionally does what he knows he ought not to do, e.g. there is no reason to believe that there was a time when Oedipus chose to do what he thought he should not choose to do, for the opacity of what is chosen allows 'he chose to sleep with Jocasta' to be consistent with 'he did not choose

to sleep with his mother'. The revaluations I have in mind are relatively simple; one learns that what one chose to do was only a partial description of what occurred, that what actually occurred was regrettable, and so, if this is applicable, one resolves to see to it that an action of that sort does not recur. Turning to an undramatic choice in the market, it is possible that in voluntarily giving Chamberlain a quarter for the opportunity of watching a game, a spectator involuntarily becomes a participant in a social move, D_1 to D_2, unknowingly helping to bring about a distribution which that spectator detests. For this man, it is not true that he '. . . voluntarily moved from it [D_1, the just distribution] to D_2'; so he can naturally regret at D_2 having played his petty part. He can simply realize that his quarter transfer was a mistake on grounds of justice. If one insists that this spectator cannot complain about the unfortunate consequences of his action because his action was voluntary and so those consequences were brought about voluntarily, if one slides with Nozick from 'each of these persons *chose* to give twenty-five cents of their money to Chamberlain' to '. . . people voluntarily moved from it [D_1] to D_2', one must be ready to believe that any consequence of any conjunction of voluntary acts is something people voluntarily bring about, e.g. that there are (are not) an even number of cars on the freeways of Los Angeles. There is no reason to slide with Nozick. Thus, our imagined spectator might be unable to complain, for example, when he transfers his quarter on the first day of the season, but he can complain when at the end of the season he learns that Chamberlain received $250,000.

III

If someone can legitimately find fault with the result of a series of choices on grounds of justice, there is no reason to believe that he must criticize one of the choices as unjust. Each participant in Nozick's Wilt Chamberlain example was '. . . led by an invisible hand to promote an end which was no part of his intention'. It is possible that this end is one which each spectator would have liked to avoid, i.e. an end which each would have sought to avoid in conditions of optimal information. The unforeseen end could be undesirable, so let us suppose it is undesirable; then each spectator who learns that his choice promoted this end would agree that his petty exchange with Chamberlain did not fully promote his ideals. To see how the invisible hand might fail to preserve the specific ideal of justice, assume that the stock socialist forecasts as to how substantial private wealth will be spent are true. Investments gaining short-term profits for the wealthy investor will be preferred to investments gaining long-term collective benefits: mansions before public libraries, etc. With this acknowledged, those who brought about the move from D_1 to D_2 might come to realize at the end of the

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. **Articles submitted for publication should be addressed to Christopher Kirwan, Exeter College, Oxford OX1 3DP.** Contributors are asked to note the following requirements.

Articles should normally not exceed 3,000 words in length. **Occasionally,** however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons of the requisite value,** whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

1977

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

ANALYSIS 38.1

EDITORIAL

Publication of this first issue of volume 38 has been deferred from October to January. Subsequent issues in the volume will appear in March, June and October, and it is planned that future volumes shall each fall, like this one, within a single calendar year.

THE IRREDUCIBILITY OF EVENTS

By Irving Thalberg

I. The Stakes

WHAT is an event? The question is odd but not idle. Some live philosophical debates turn upon the notion of an event. The mind-brain identity fracas is a prime example. Materialists and their more or less dualistic opponents are not concerned about whether some elusive object, our mind, could be identical with our neural apparatus. Rather, materialists contend, and their critics deny, that our mental states and activities could be the very same events as certain electro-chemical happenings within our brain. Another kind of identity puzzle comes up in action theory. We may safely assume that a human deed is an event, and has other events as its effects. But what species of occurrence is it? Normally our bodies, or parts of them, move when we act. Could our performance itself be the same event as this contemporaneous motion of our body? Some metaphysicians would insist that for an action to occur, there must be something more than agitation of our limbs: perhaps also an inner thrust of our will—a volition, a covert episode of 'bringing about'—which somehow makes our limbs move. But then what sort of event is an exercise of will or a 'making happen'?

I do not intend to enter, much less to settle, these and other perplexing debates. I evoked them only to illustrate why it is not frivolous to inquire what events are. The remaining two sections of this essay take up the most imaginative answers that contemporary metaphysicians have propounded. With one clear exception, these answers sound reductive. For most of the writers I consider attempt to define our concept of an event solely by reference to notions of some different, presumably more fundamental, and perhaps more familiar category. Despite their ingenuity, I find these attempts unsuccessful—for reasons I shall catalogue.

I cannot prove that no deflationary analysis will ever work; but that prognosis should gain support from my reasoning here.

Broadly speaking, my animadversions toward reductive theories of events go counter to one of Professor Strawson's most appealing contentions. Strawson holds that in our everyday conceptual scheme, what he calls 'material bodies' possess a kind of 'ontological priority' over all other categories of individual things—including particular occurrences (1959, esp. pp. 16–57). Of course, to say that events take an ontological back seat to material objects is hardly to assert that we can define events in predominantly material-object terms. Nevertheless, Strawson's 'ontological priority' claim encourages reductive ambitions (see Goldman, 1971, p. 773, discussed later).

In briefest outline: speech situations appear to furnish Strawson his most important criterion for deciding which of two categories of particular is ontologically nearer rock-bottom than the other. According to Strawson's lexicon, a language-user 'identifies' or 're-identifies' a particular of some type when she or he manages to single out the particular by naming it, by describing it with sufficient precision, by engaging in an act of perceptual discrimination—even by grabbing it. Strawson's assumption is that if speakers can pinpoint items of type X for their listeners without having already identified any specimens of type Y, then particulars of type X rank as ontologically prior to those of type Y.

Professor Moravcsik has persuasively challenged Strawson's claim of prior identifiability, and hence ontological pre-eminence, for material bodies. Moravcsik insists that we can only re-identify a particular knife, table, animal or person if we also single out some events in which it has figured. Moravcsik explains:

> we have to locate two segments of time which the body allegedly occupied and occupies . . . [We need] reference to events, or times (*e.g.*, 'The knife in front of me now is the same knife with which I cut an orange an hour ago' . . .) . . . Furthermore, reference to times presupposes some reference to events.
>   . . . [T]hough part of the evidence for a claim of re-identification may be qualitative sameness (*e.g.*, 'The knife looks the same'), the statements of evidence would have to connect this with events of observation (*e.g.*, 'The knife looks the same as the one I . . . saw an hour ago') . . . [T]hough the identification of [events] may depend . . . on the identification of material bodies, this latter dependency makes the relation simply a symmetrical one . . . (1965, p. 116).

Professor Davidson also endorses a non-Strawsonian 'ontological parity' viewpoint. Davidson believes

> there is . . . a symmetrical dependence of the category of objects on the category of events . . .
>   . . . Substances owe their special importance in . . . identification to the fact that they survive through time. But the idea of survival is

inseparable from the idea of surviving certain sorts of change . . . [Thus] events often play an essential role in identifying a substance . . . Neither the category of substance nor the category of change is conceivable apart from the other (1970, pp. 226f.).

If I am able to show that the best-known schemes for reductively analysing events in material-body and other terms are dubious, then that would strengthen the position of 'symmetry' ontologists like Moravcsik and Davidson. For if we fail to eliminate the category of events, why should we consider it subordinate to the category of physical objects?

Now one final bit of stage setting. For my broader and my narrower purposes, I doubt that we need draw special distinctions within the generic category of events. We can deploy as examples steady or intermittent processes; sudden or gradual transformations; enduring or transitory states; phases, stages, conditions, situations, circumstances, dispositions, beginnings or endings. We may be equally free with the objects—'material' or otherwise— that take part in events.

II PROPOSITIONAL THEORIES

Professor Chisholm is the most thoroughgoing defender of this approach (see also Wilson, 1974). His account, unlike the 'property-exemplification' doctrines I review in section III, seems not to be reductive. I think his aim is to simplify and systematize, rather than to replace, our event language. Overall, his view is that for an event to occur is for a 'state of affairs' to 'obtain' (1970, pp. 15ff.). A 'state of affairs' seems proposition-like to me because Chisholm assigns it the jobs which most philosophical friends of propositions reserve for such alleged entities. Chisholm says that a state of affairs can be 'the object of belief, or of hope, or of wonderment, or of any . . . other intentional attitudes'; also it is something to which 'the laws of propositional logic may be interpreted as being applicable' (1970, p. 19). Instead of delineating states of affairs in straightforwardly propositional form, by means of a sentential clause, Chisholm prefers the gerundive style—for example, when he speaks of 'that state of affairs which is John sitting' (1970, p. 20). Here is the fullest statement of his outlook:

> For every well-formed sentence, there is a corresponding gerundive . . . Thus for 'Socrates is mortal' there is 'Socrates being mortal' . . . The gerundives of well-formed sentences may be said to designate states of affairs.
>   . . . States of affairs are of two sorts—those that obtain and those that do not obtain. Hence if a well-formed sentence is true, then the state of affairs . . . is one that obtains . . . And so we may say that there *are* states of affairs, some of which obtain and some of which do not obtain. In place of 'obtains', we may say 'takes place', or 'occurs', or 'is actual', or even

'exists' (but if we use 'exists', we should say 'There *are* states of affairs that do nòt exist' and not 'There *exist* states of affairs that do not exist' (1971, pp. 39f.).

Just as, according to devotees of propositions, there *is* the proposition that John is sitting, regardless of John's posture, Chisholm would hold that there *is* 'that state of affairs which is John sitting', no matter what goes on. John might be standing on his head, squatting, cartwheeling, prone, or ensconced in his favourite armchair; in all these situations, the state of affairs *is*.

You may puzzle over the verb 'obtain', which Chisholm affixes to states of affairs. In English this verb means 'get hold of by effort'. Since only animate creatures have goals, and put forth effort to attain them, it sounds odd if we use the verb 'obtain' to characterize states of affairs. More important, 'obtain' is a transitive verb, requiring completion by a direct object phrase. 'John obtained' is grammatically incomplete. What did John obtain—beer, lumber, a building permit? 'That state of affairs which is John sitting obtained' lacks a direct object phrase, and contains an inadmissible subject term. For how can a state of affairs expend energy pursuing an aim?

We can avoid these conundrums if we put some of Chisholm's alternative verb phrases, which do not require direct objects or animate subjects, in the place of 'obtain'. But then we face a deeper difficulty: to what extent has this kind of theory elucidated our notion of an occurrence? What metaphysical insights does it give us into the nature of events? Suppose we agree that there are 'two sorts' of states of affairs, by analogy with the way propositions are either true or false. Next we interpret any of the verbs we have chosen to mark this dichotomy—'obtains', 'takes place', 'is actual', 'exists' and their antonyms—as literally as possible. We seem to be describing something that happens, or perhaps fails to happen, to some quasi-propositional entity. This sounds suspiciously event-like. We appear to have elucidated events generally—including the kind of event in which a physical object, or human being like John, has a role—by reference to one special kind of happening which only features states of affairs. But recall that we were uncertain what events are—what it is for John to be sitting. How could it then enlighten us if we are told that this episode in John's career is merely an 'obtaining', or 'taking place' or whatever event in the life of 'that state of affairs which is John sitting'?

Presumably the 'state of affairs' doctrine is an attempt to reconstruct rationally our thinking about events, and not an attempt to eliminate concepts of the event-family. Does it in any way simplify or systematize our thinking? For every standard happening—such as the blizzard which paralysed Chicago in February 1967—we have a new one: here, the 'occurrence' or whatever of a stormy state of affairs. So we have not

economized in the domain of events to compensate for our mintage of state-of-affairish entities. Moreover, we have incurred a double loss of simplicity. These extra state-of-affairish items, and their high jinks, are mysterious in a way that old-style events and their down-to-earth participants are not. We can observe, film, tape-record, overlook, misperceive, sometimes control, sometimes lose control of, ordinary events, people and 'material bodies'. We can lay our hands on and examine a person or a 'material body'. Do we have any comparable forms of access to states of affairs, or to the occurrings they undergo? Pending detailed clarification of such points, we have no reason to concede that the 'states of affairs' analysis has simplified things, by disclosing the 'basic' events behind the phenomenal order.

A state-of-affairist might object that I have misunderstood Chisholm's variant of this approach. The complaint might be that a states of affairs theory does not have to be an attempt to eliminate, or to reconstruct, our event discourse. Instead, it can be a formal, axiomatic calculus, which happens to contain verb-like symbols resembling the English words 'take place', 'occur', and so on. Hence I am mistaken to assume that a state-of-affairist endows these terms of his calculus with any of their normal meaning.

In response to this formalistic manoeuvre, I suppose I would meekly withdraw my challenges. But instead I would ask: How can an uninterpreted, purely formal system of notation illuminate the nature of events?

If we put aside the exegetical question of how Chisholm intended his particular states of affairs analysis of events, I think we can draw three hypothetical conclusions so far. (i) If we take this kind of theory to be reductive, we have to count it a failure, since it leaves us with an event: the 'occurring' or 'taking place' of a state of affairs. (ii) If we consider it as a revelation of the bedrock events, and the fundamental entities which participate in them, we must reject it for its lack of conceptual economy. (iii) If we look upon it as a calculus, we must concede that it is irrelevant to our metaphysical quest.

III. PROPERTY-EXEMPLIFICATION THEORIES

This time my conclusions will not be hypothetical, because at least one of the analyses I shall probe is explicitly reductive. I am alluding to Professor Goldman's view of those events which are human actions. Goldman expects his theory to 'lay bare the nature, or ontological status, of an act' (1971, p. 768). His ambitions extend to other events besides actions. Invoking Strawson's thesis, which I mentioned earlier, that 'material bodies' have an ontological head-start on other particulars, Goldman announces:

Instead of treating actions (or events) as a primitive or irreducible cate-
gory, our account reduces act tokens [particular acts] to persons, act-
properties, and times. This supports the Aristotelian-Strawsonian
ontology in which substances are primary, and events . . . derivative
(1971, p. 773).

Now state-of-affairists and other propositional theorizers only have
their monolithic entities, which take place or whatever *en bloc*. Such
metaphysicians cannot reduce events to the 'substances' which are in-
volved in a state of affairs. On the kind of theory Goldman advocates,
however, happenings are more directly tied to the substances participat-
ing in them. To illustrate, here is Goldman's deflationary account of how
a homeowner, John, figures in an unspectacular horticultural incident.
Goldman writes:

When we ascribe an act to an agent, we say that the agent exemplified an
act-property (at a certain time). When we say . . . 'John mowed his lawn',
we assert that John exemplified the property of mowing his lawn . . .
. . . A particular act, then, consists in the exemplifying of an act-
property by an agent at a particular time (1970, p. 10).

This pattern of analysis seems to have originated with Professor
Kim, who applied it to every type of occurrence. Kim's first proposal
was to equate 'event or state' with 'a particular (substance) having a
certain property, or more generally a certain number of particulars
standing in a certain relationship to one another' (1966, p. 231). In a
subsequent essay, co-authored with Professor Brandt, Kim supplanted
'having' by another verb, and 'particulars' by the curious substantive
'location'. They tell us:

To say that there is an event . . . is to say that some logically contingent
property (set of properties) is instantiated at a specific time and 'location'
. . . '[L]ocation' . . . [is] a technical term which may be, but need not be,
construed to refer to physical position . . . [I]f we take . . . Socrates as one
of the fundamental individuals of the world, we could construe the
'location' of Wisdom as just being 'of Socrates' (1967, p. 516).

In his longest discussion, Kim substitutes 'exemplify' for 'instantiate',
and discards the bewildering concept of a 'location'. He declares an event
to be

a concrete object (or n-tuple of objects) exemplifying a property (or n-adic
relation) at a time . . . Events . . . have something like a propositional
structure; . . . the exemplification of property P by an object x at time t
bears a structural similarity to the sentence 'x has P at t'.
. . . Linguistically, we can think of '$[(x_n, t), P]$' as the gerundive
nominalization of . . . 'x_n has P at t'. Thus '[(Socrates, t), drinks hemlock]'

can be read 'Socrates' drinking hemlock at t'. Notice that $[(x, t), P]$ is not the ordered triple . . . x, t, and P; the triple exists if x, t, and P exist; the event $[(x, t), P]$ exists only if x has P at t (1973, pp. 222f.).

I have quoted Goldman and Kim at length in order to highlight the richness and ingenuity, as well as the similarity, of their analyses. One difference is that Kim denies that his property-exemplification theory is ' "eliminative" or "reductive" '—although he hopes it will 'tell us something about the metaphysical nature of events' (1976, p. 162). Indeed, Kim appears to believe that his account uncovers something fundamental. He declares: 'events are, essentially, structured complexes of the sort the theory says they are' (1976, p. 173). Thus both Kim and Goldman flatter our reductionistic yearnings. But two straightforward questions ought to make us hesitate—if we are as yet uncommitted to the Kim-Goldman *Weltanschauung*.

The first question was anticipated but not elaborated by Moravcsik in the article I cite in section I. He was discussing Strawson's notion that we re-identify a 'material body' as one that we had identified at a previous time. Moravcsik's comment was that 'reference to times presupposes some reference to events'. By way of expanding this, I would ask would-be reductionists how they can hope to expunge talk of events from their analysis of events. If they hold that an event is 'a concrete object . . . exemplifying a property . . . at a time', doesn't their final clause hide 'some reference to events'? Surely times are times at or during which this or that goes on! Every dating system which is used to specify moments and intervals of time does appear to make explicit reference to such happenings as the periodic appearance of sunlight, the change of direction of the shadows it casts, and its disappearance. Calendars specify times like days of the month, months of the year, and so on, by reference to the positions and movements of planets. At the very least, deflationary property-exemplification analysts owe us a story of their 'times' which is demonstrably free of reference to astronomical and other happenings. Until then, we can dismiss their purported reduction as incomplete. The same could be said against any state-of-affairist who imagined that he or she had eliminated events from that analysis. Whatever it is for a state of affairs to 'take place', it must also go on 'at a time'.

This brings me to my second question for property-exemplifiers. It is a variant of my suspicion that state-of-affairists led us in a circle when they told us that an event consists in a state of affairs taking place. Now we recall Kim's remark that an event is more than 'the ordered triple . . . x, t, and P'. Only a verb can knit these items into a happening. Property-exemplifiers cagily spurn the giveaway event-verbs used by state-of-affairists: 'occur', 'take place'. These would be ungrammatical here anyway. But whatever neutral-sounding verb property-exemplifiers recruit to bring vitality to object, property and time, we should have a

question ready. We should ask: Isn't this 'having', this 'being', this 'instantiating' or 'exemplifying' between object and attribute some kind of event—a transaction or getting-together among them 'at a time'?

I have never fathomed the ontology of unattached properties, plus as yet featureless 'objects' which by themselves do and undergo nothing. So evidently I cannot say what kind of happening I expect an 'instantiating' or 'exemplifying' to be. But the prepositional phrase 'at a time' is a sure-fire clue that the encounter of property and object is an event. This phrase is glaringly inappropriate when we talk about relationships which are not events. For instance, wouldn't it be odd to assert that five plus four equals nine at midnight, April 2, 1977—or at all times? The mathematician's verb 'to equal' does not mark an occurrence, or even a persisting state. Thus if Kim and Goldman wanted the 'exemplifying' relationship to be uneventful, they would not bring in the tell-tale modifying expression 'at a time'. Whatever unassigned properties and the quality-less objects they are destined for may be like, if their joining up bears a date or a duration, then I suspect an event has been smuggled in.

Suppose that property-exemplifiers are miraculously persuaded to renounce their deflationary enterprise. Then, although they cannot pretend to have reduced events to something else, they might still claim that they have boiled down the whole confusing array of goings-on to a single, elementary kind: the displaying of a quality by an object. If they say this, my new question would be: In what sense does an episode of 'exemplifying' underlie all other events? How are they 'essentially' —as Kim puts it—dynamically 'structured complexes' of this sort? How would it 'lay bare the nature, or ontological status' of such events as the 1967 Chicago blizzard, and Socrates' ingurgitation of hemlock, if we learn that these are really cases of an object exemplifying a property?

One advantage of the Kim-Goldman analysis over propositional or state-of-affairs theories is that it spotlights the people and objects that participate in an event. These, at least, can be singled out for study. But the properties that Kim and Goldman saddle us with seem just as elusive, as inaccessible to scientific and everyday observation, as were states of affairs. So I doubt that any boiling down, or other conceptual *cum* ontological economy would be achieved if we imagine that underneath familiar happenigs there are objects and properties getting together.

If we reject the property-exemplifier's simplicity claim along with his reductive thesis, he can fall back upon formalism, as a state-of-affairist might. But again, there is a rub. His axiomatic system may be elegant, and intimidating to the non-mathematician. However, if we cannot interpret its key symbols, such as 'exemplify', what can it teach us about events? I think events will have to remain a primitive category.

The resourceful attempts of state-of-affairists and property-exemplifiers to analyse the category away helps us understand why it cannot be displaced.[1]

University of Illinois at Chicago Circle   © IRVING THALBERG 1978

[1] I am grateful to a referee from ANALYSIS for searching criticisms of my previous draft. I have tried to take account of them here.

REFERENCES

Brandt, R., and Kim, J. (1967) 'The Logic of the Identity Theory', *Journal of Philosophy* (henceforth '*JP*'), XIV, pp. 515–37.
Chisholm, R. M. (1970) 'Events and Propositions', *Nous*, IV, pp. 15–24.
——(1971) 'On the Logic of Intentional Action', in R. Binkley *et al.* (eds.), *Agent, Action and Reason*. Toronto: Univ. Press, pp. 38–69.
Davidson, D. (1970) 'The Individuation of Events', in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*. Dordrecht: Reidel, pp. 216–34.
Goldman, A. (1970) *A Theory of Human Action*. Englewood Cliffs: Prentice Hall.
——(1971) 'The Individuation of Action', *JP*, LXVIII, pp. 761–74.
Kim, J. (1966) 'On the Psycho-Physical Identity Theory', *American Philosophical Quarterly*, III, pp. 227–35.
——(1973) 'Causation, Nomic Subsumption, and the Concept of Event', *JP*, LXX, pp, 217–36.
——(1976) 'Events as Property Exemplifications', in M. Brand and D. Walton (eds.), *Action Theory*. Dordrecht: Reidel, pp. 159–77.
Moravcsik, J. (1965) 'Strawson and Ontological Priority', in R. Butler (ed.), *Analytical Philosophy*, 2nd Series. Oxford: Blackwell, pp. 106–19.
Strawson, P. F. (1959) *Individuals*, London: Methuen.
Wilson, N. (1974) 'Facts, Events and Their Identity Conditions', *Philosophical Studies*, XXV, pp. 303–21.

FREGE, DUMMETT AND THE PHILISTINES

By NICHOLAS MEASOR

THE claim that paradoxical consequences follow from Frege's thesis that incomplete expressions such as ' . . . is fat' refer to concepts is too well known to need extensive exposition. There seems to be a strong *prima facie* case for saying that the predicate '. . . . is a concept' cannot be truly applied. For '. . . . is a concept' appears to be an incomplete expression of the first level, i.e. one into whose argument place one would insert names. So on Fregean principles the predicate can only be applied to objects, not to concepts (objects being the referents of names) and thus can only be applied falsely. Another way of expressing the difficulty is to point out that if one tries to apply '. . . is a concept' to e.g. the referent of '. . . is fat' one will first have to devise a name for that referent to insert into the argument place of '. . . is a concept'. But once one tries to apply the name to the concept one will, as Frege would put it (*Philosophical Writings of Gottlob Frege*, translated by Peter Geach and Max Black, Oxford 1960, p. 197), convert the concept into an object or 'represent' the concept by an object.

For those with a taste for ontology a particularly crucial consequence of this problem is that it will follow that it is false to say that there exist any concepts. For if '. . . is a concept' is an incomplete expression of the first level 'there exists something which is a concept' would, if true, state the existence of the members of some set of objects, not concepts—a *reductio ad absurdum* of the idea that concepts exist. But the suggestion that incomplete expressions refer to concepts is cut down to size somewhat if there do not exist any concepts, while the claim that '. . . is fat' refers to a concept would hardly be strengthened by the fact that there was no such thing as the concept to which it referred.

Michael Dummett has offered an ingenious treatment of the topic which tries to rehabilitate Frege's theory by disarming the objection (Michael Dummett, *Frege: Philosophy of Language*, London 1973, pp. 213–218; all references to Dummett in this article will be to this work). Dummett's proposal is that we should abandon the apparent name 'the concept which ". . . is fat" refers to' as a means of referring to the appropriate concept, and use instead the first-level incomplete expression 'what ". . . is fat" stands for', where the 'what' is predicative as in 'Blue Peter is what Tsarina and Celerity both are' (namely a racehorse). Similarly instead of applying the apparently first-level '. . . is a concept' to concepts we should use the second-level 'everything either is or is not . . .' as a predicate which can be applied truly to all (and only) first-level concepts. Thus 'everything either is or is not what ". . . is fat" refers to'

will succeed where 'the concept which ". . . is fat" refers to is a concept' failed, providing an obvious truism rather than an inevitable falsehood.

There is a certain presumption that Dummett's suggestion is Fregean. For Dummett claims to have seen a paper containing it in Frege's *Nachlass,* together with an editorial letter rejecting the paper for publication (the paper in question does not, however, appear in Frege's published *Nachlass*). Geach, commenting approvingly on Dummett's treatment, describes the (reported) rejection of the article as a triumph for the Philistines (*Mind,* Vol. LXXXV (1976) p. 438). I propose here to strike a blow for Philistinism by showing that Dummett has not succeeded in revivifying the Fregean thesis.

Even those of us who are Philistines must concede, I think, that Dummett has succeeded in showing that some meaning could be attached to the word 'refer' which would make it possible to insert an expression into the second gap in ' ". . . is fat" refers to . . .' so as to produce a true sentence. But, if we cast Dummett in the role of a surrogate Samson acting on Frege's behalf, the concession is far from being an admission that he has succeeded in pulling down the temple of Dagon.

It is clear that reference as construed in Dummett's account of reference to concepts is a different sort of device, perhaps even reference in a different sense of 'reference', from the device one uses when naming. For '*W* refers to . . .' is on Dummett's account a first-level predicate when *W* is a name, but a second-level one when *W* is a first-level predicate. Indeed Dummett allows that his two strands in the ascription of reference, reference as the name/bearer relation and reference as semantic role, become dissociated when the reference is to concepts, only the semantic role remaining (e.g. pp. 245, 523–524). Something more is needed if Dummett's doctrine is to have real philosophical bite, and that something is a justification of the thesis that concepts, as resuscitated by Dummett, *exist*. That part of Dummett's theory which we have so far examined may defend the claim that incomplete expressions refer, but it does not show that their referents exist. As I have already remarked, however, the thesis that incomplete expressions refer to concepts amounts to little if concepts do not exist.

In connection with the ideas which are prominent in the last paragraph it is worth mentioning the marked similarity between Dummett's views and those expressed by Montgomery Furth in 'Two Types of Denotation' (*Studies in Logical Theory,* American Philosophical Quarterly Monograph No. 2, ed. Nicholas Rescher, Oxford 1968, pp. 9–45): while Dummett implicitly allows that there is more than one form of reference, Furth explicitly states that there is; Furth's analyses of ascriptions of reference to predicates are not dissimilar from Dummett's (cf. Part III of Furth's article); and their approaches are akin in that they both lay emphasis on the importance of the tendency to introduce quantification

in connection with incomplete expressions (cf. Furth p. 28). Indeed in one respect Furth gives an extra twist to the argument. For, to put the point in Dummettian terminology, he not only claims, like Dummett, that incomplete expressions have reference in the sense of semantic role (although not in the name/bearer sense), he also puts forward arguments not, I think, found in Dummett to demonstrate that the semantic roles of names and predicates are strikingly parallel (Furth pp. 31–40).

Nevertheless, despite this extra dimension to Furth's article, it appears to me that he would fare no better than Dummett in the face of the objections which I raise. For even if it is conceded that the semantic role of incomplete expressions is to denote (in a fashion different from that of names), it by no means follows that the denotations of these expressions *exist*. If the difficulty which I shall discuss afflicts Dummett's account of the existence of concepts, it afflicts Furth's equally. Indeed while there is a non-Dummettian dimension to Furth's treatment, there is also a non-Furthian, and commendable, ingredient in Dummett's observations, namely the ingenious defence of the ascription of existence to concepts which I discuss in the last eight paragraphs of this article.

Let us return to our main theme. Dummett does indeed claim that concepts exist (e.g. p. 218 and p. 245) and he suggests that the use of the name/bearer relation as the prototype of reference is justified exactly to this extent, that the explanation of the semantic role of an expression will always consist in the association of some extralinguistic *entity* with it (p. 524).

The crucial issue, therefore, is whether Dummett's account of existential statements about concepts succeeds. His official account of the matter (p. 218) provides an analysis of singular existential statements about particular concepts. The existential statement about the concept which '. . . is a philosopher' stands for, for example, should be understood as '(∃f)(x)(fx if, and only if, x is a philosopher)'. The rationale of this analysis is supposed to be that the coextensiveness of two predicates provides a relationship between the concepts to which they refer analogous, *mutatis mutandis,* to the relationship which holds between identical objects. '(∃f)(x)(fx if, and only if, x is a philosopher)', it is claimed, performs the same function for the concept which '. . . is a philosopher' refers to as that which '(∃x) x is identical with Mount Everest' does for Mount Everest (which function is often thought to be that of ascribing existence to Mount Everest).

Although Dummett does not specifically explain how the general statement that the referents of incomplete expressions exist is to be paraphrased, it is clear that it should be possible to extend his theory so as to provide such a paraphrase. According to him the predicate truly applicable to all, and only, concepts is the second-level 'everything either is or is not . . .'. So what is needed is some *third-level* predicate into whose

argument place we can fill the second-level predicate so as to express
the existence of concepts in the same way that the first-level 'is an object'
can be inserted into 'there exists something which . . .' so as to assert the
existence of objects.

It is easy to see what the Dummettian candidate to serve as the ap-
propriate third-level predicate would be. He says that the fact that our
language contains higher-order idiomatic quantification guarantees that
we have a conception of the existence of concepts (pp. 223, 245, 408).
We should expect, therefore, that higher-order quantification will play a
role in existential statements about concepts comparable to that played
by first-order quantification in existential statements about objects. So
the appropriate third-level predicate should be the predicative 'some-
thing' (as in 'he is something which I am not, i.e. kind') and the Dum-
mettian assertion that concepts exist will come out as the assertion that
everything either is or is not something.

It is not unreasonable to suggest that '($\exists$f)(x)(fx if, and only if, x is a
philosopher)' and 'everything either is or is not something' are as near as
the Frege/Dummett theory can come to ascribing existence to concepts
successfully. But they are not near enough.

What could be meant by the claim that Dummett has shown that
concepts exist? Only, surely, that 'concepts exist' and 'there exist con-
cepts' mean the same thing as 'everything either is or is not something',
and that the statement that the appropriate concept exists means the
same thing as '($\exists$f)(x)(fx if, and only if, x is a philosopher)' and that all
these statements are true. But there are two criticisms of this proposal
which reduce the appeal of Dummett's stratagem.

In the first place even if 'concepts exist' means the same thing as
'everything either is or is not something' it does not follow that either
statement ascribes existence to the things which, on the Fregean account,
incomplete expressions refer to. For 'there exist concepts' has every
appearance of ascribing existence to a range of Fregean objects, since it
is an abbreviated version of the sentence which we form by filling the
first-level '. . . is a concept' into the argument place of the second-level
'there exists something which . . .'. But if the statement ascribes exis-
tence to a range of Fregean objects it can hardly be ascribing existence to
the things which incomplete expressions, according to Frege, refer to.
Even if 'there exist concepts' means the same as 'everything either is or
is not something' this shows at most only that there exist a range of
objects called 'concepts' and that the statement that they exist is justified
by the fact that it is paraphraseable by a true statement saying something
different about the things to which 'everything either is or is not . . .'
applies.

Against this argument Dummett might counter that '. . . is a con-
cept' is not truly a first-level expression predicable of objects since it is

merely a *façon de parler* eliminable in favour of the second-level 'everything either is or is not . . .' (an argument in the same spirit as his proposal that 'wisdom' is not truly a name—see pp. 78 and 255–256). But whether or not it is so eliminable, the syntactic properties of '. . . is a concept' are sufficient for our Philistine purposes. For even if the sentences 'there exists something which is a concept' and 'everything either is or is not something' did mean the same thing this would not necessarily show that the two sentences ascribed existence to the things which the predicate 'everything either is or is not . . .' applied to. In fact it would only show this if the predicate and '. . . is a concept' had the same extension. But they cannot have the same extension since they are never substitutable for each other *salva veritate* (a reflection of their differing syntactical properties).

A second difficulty with the idea that Dummett has shown that the referents of incomplete expressions exist is that the propriety of using the word 'exist' in 'concepts exist' (as interpreted by Dummett) can be called in question. Admittedly anyone has the right to introduce a sentence 'concepts exist' and to stipulate that it should mean 'everything either is or is not something'. And anyone has the right to stipulate that a singular statement containing the word 'exist' should mean '($\exists$f)(x)(fx if, and only if, x is a philosopher)'. It is quite another matter, however, to suggest that anyone can properly imply that the sense of 'exist(s)' introduced in this way has anything to do with the sense or senses of 'exist' which are commonly used. But someone who claimed that Dummett has shown that concepts exist would appear to be claiming this— wrongly, I suggest. The consequent sense of 'exist' is purely Dummettian.

These objections are, I think, sufficient to show that the Dummett/ Frege thesis does not safeguard the ontological status of concepts from the onslaught of the Philistines. But there is one further suggestion of Dummett's which we must consider since, if true, it would forestall my criticisms.

We are familiar with the first-level expression of natural language 'there is such a thing as . . .'. We can fill a name into the gap in this expression so as to express the existence of the thing referred to by that name. Dummett claims (p. 218) that there is a *second-level* expression of natural language which is also recognizably existential in the sense that it clearly ascribes existence to the referent of the first-level predicate filled into the gap. Consider, for example, the following sentence: 'there is such a thing as being a philosopher'. This sentence certainly looks existential, and it is true. It is Dummett's contention that it truly ascribes existence to the concept which '. . . is a philosopher' refers to. For he asserts that 'being a philosopher' is here to be construed as a first-level incomplete expression, not a name, and hence that 'there is such a thing

as . . .' is here second-level. The claim that 'being a philosopher' is a first-level predicate rather than a name is supported by the idea (p. 216) that it can occur at the beginning of sentences where it is followed by a non-copulative 'is' and then another predicate—sentences like 'under-paid is what Peter does not want to be' and 'being laughed at is what Henry most dislikes' (the examples are Dummett's own).

If Dummett's account of 'there is such a thing as being a philosopher' is correct then both the objections to his theory expounded above will probably be frustrated. On the one hand if there is an existential natural language version of the statement ascribing existence to some particular concept then it no doubt follows that the Dummettian statement that concepts exist does not use 'exist' in an idiosyncratic manner. And, on the other hand, if 'there is such a thing as . . .' is in this sentence second-level, not first-level, then we are not faced with the difficulty that it ascribes existence to an object rather than a concept.

But whatever 'there is such a thing as being a philosopher' truly ascribes existence to, it cannot truly ascribe existence to the concept which '. . . is a philosopher' refers to.

If it did truly ascribe existence to this concept then in the case of any other predicate which refers to a concept it should be possible truly to ascribe existence to *that* concept by the use of the sentence formed out of the appropriate predicate in the same way that 'there is such a thing as being a philosopher' is formed out of '. . . is a philosopher'.

Now according to the Fregean theory *all* predicates, including those with a self-contradictory air about them such as '. . . is in two places at once', make reference to concepts (Dummett p. 219). Therefore if 'there is such a thing as being a philosopher' truly ascribes existence to the concept which '. . . is a philosopher' refers to then 'there is such a thing as being in two places at once' must truly ascribe existence to the concept which '. . . is in two places at once' refers to. But while '$(\exists f)(x)(fx$ if, and only if, x is in two places at once)', the formal sentence which, according to Dummett, ascribes existence to this concept, is true, all our intuitions about natural language tell us that 'there is such a thing as being in two places at once' is false. There is *no* such thing as being in two places at once. So 'there is such a thing as being in two places at once' cannot be used truly to ascribe existence to the Fregean referent of ' . . . is in two places at once'. And from this it follows (by *modus tollens* using the previous paragraph as a premiss) that 'there is such a thing as being a philosopher' cannot ascribe existence to the Fregean referent of '. . . is a philosopher'.

I must admit that Dummett is aware of the difficulty about self-contradictory predicates and believes, rather injudiciously I suspect, that it can be reconciled with his theory: 'There is, indeed, a divergence here between Frege's criterion for the existence of a concept and that which

underlies the use of second-level generalization in natural language'
(p. 219). He appears to think that there is a single set of entities, namely
concepts, such that the users of natural language have one set of existence
criteria for those entities and the users of the formal language in which
we find e.g. '(∃f)(x)(fx if, and only if, x is a philosopher)' have another set
of existence criteria. In consequence he is prepared to assert that
'(∃f)(x)(fx if, and only if, x is in two places at once)' can be used truly to
ascribe existence to the appropriate concept even though 'there is such a
thing as being in two places at once' cannot be so used. But this position
is deeply suspect.

The situation is this. We have two sets of sentences, one of sentences
of the same form as 'there is such a thing as being a philosopher', the
other of sentences of the same form as '(∃f)(x)(fx if, and only if, x is a
philosopher)'. The reason given for supposing that the second set of
sentences is existential is that the first set is existential and the second set
paraphrases the first. What should we infer, therefore, when we discover
that there is a divergence in truth-conditions between some members of
the second set and the corresponding members of the first set? Dummett
suggests that we should infer that the users of the second set and the
users of the first have adopted different criteria for the existence of
the entities in question. But surely a more prudent inference is that the
sentences of the second set have nothing to do with existence at all, and
that their users therefore have no existence criteria in mind. For with the
divergence in truth-conditions between the two sets of sentences we lose
the one reason for supposing that the sentences of the second set are
existential, namely the supposed fact that the second set paraphrases the
first. According to Dummett the divergence in truth-conditions pro-
vides a reason for distinguishing the natural language criterion for the
existence of concepts from the Fregean criterion. In fact, however, the
divergence demonstrates that there is nothing which can properly be
called a Fregean criterion for the existence of concepts.

We can conclude, therefore, *contra* Dummett, that, even if he has
shown that incomplete expressions in some sense refer, he has not shown
that their referents exist. There is nothing Philistine, in the pejorative
sense, in objecting that these resuscitated Fregean concepts are too
shadowy to cause alarm to those whose semantic theory does not in-
corporate them.

University of Leicester   © NICHOLAS MEASOR 1978

HOW TO MAKE A NEWCOMB CHOICE

By Don Locke

I. Newcomb's Problem

I am given two sealed boxes and two choices, labelled not as is standard in the literature but rather, for obvious mnemonic reasons,

Choice One: to take one box only, Box One;

Choice Two: to take the two boxes.

I know that Box Two contains a thousand pounds but I do not know how much, if anything, is in Box One. What I do know is that an extraordinarily able Predictor, whose predictions have invariably been confirmed in a multitude of such cases, has predicted which choice I will make, and has acted on his prediction as follows:

if he predicts Choice One he rewards my moderation (or faith in his powers) by placing a million pounds in Box One;

if he predicts Choice Two he penalizes my greed (or lack of faith) by leaving Box One empty.

Which choice should I make? At first sight it seems obvious that I should make Choice One. Evidently I can be sure that whichever I choose the Predictor will have predicted that choice, and hence that if I open Box One alone I will find a million pounds inside, whereas if I open both boxes I will find Box One empty, and so gain only the thousand I know to be in Box Two. But on second thoughts it seems equally obvious that I should make Choice Two. For the Predictor has already made his choice and acted accordingly: if he has predicted Choice One and put a million in Box One then I will be better off if I make Choice Two and thus gain both the million and the thousand I know to be in Box Two; alternatively, if he has predicted Choice Two and left Box One empty I will again be better off making Choice Two and at least getting the thousand in Box Two, rather than making Choice One and getting nothing; so either way, whatever the Predictor has predicted, I will be a thousand pounds better off with Choice Two. Thus the fact that I have every reason to think that the Predictor has predicted correctly, and hence that there will be more in Box One alone if I take only Box One than there will be in both boxes if I take both boxes, gives me every reason to make Choice One; and the fact that I have every reason to think there must be more in both boxes than there will be in Box One alone gives me every reason to make Choice Two. So which *should* I choose?

Some answer in one way (e.g. R. Nozick, 'Newcomb's Problem and
Two Principles of Choice', in *Essays in Honor of Carl G. Hempel,* ed. N.
Rescher, D. Reidel 1969); some in another (e.g. M. Bar Hillel and A.
Margalit, 'Newcomb's Paradox Revisited', *British Journal for the Phi-
losophy of Science,* 1972); some in both at once (e.g. G. Schlesinger, 'The
Unpredictability of Free Choices', *British Journal for the Philosophy of
Science,* 1974). Much of the discussion has concentrated on the problem
as a problem for decision theory, to which the answer may well be that
the situation has not been sufficiently specified for decision theory to
yield an unambiguous answer (cf. I. Levi, 'Newcomb's Many Problems',
Theory and Decision 1975; J. Cargile, 'Newcomb's Paradox', *British Journal
for the Philosophy of Science* 1975). But decision theory or no decision
theory, the problem would remain, should you ever be so fortunate as to
find yourself in the situation described. Imagine yourself faced with that
choice. If it helps, imagine yourself totally ignorant of decision theory
and of the manifold alternative pieces of information which, if only you
knew of them, might resolve the problem in one way of the other. In
this state of relative ignorance, knowing only the Predictor's project and
his record, what should you choose?

II. The Arguments

The case for Choice Two seems straightforward enough: the Pre-
dictor has already laid his money down—or not, as the case may be—so
I cannot be worse off if I take both boxes; indeed either way I must be a
thousand pounds better off taking both than if I merely take Box One.
This simple point is, however, easily obscured by the puzzling suggestion
of reverse causation. But even if the actual existence of a Newcomb
Predictor were taken as strong evidence for precognition or some other
form of reverse causation, it would, presumably, be a matter of the past
prediction's being dependent on or determined by the present choice, not
a matter of the present choice's actually altering the past prediction,
making it other than it was, a suggestion which seems to me incoherent.
So even if reverse causation is involved, to the extent that I, for one, can
make sense of that notion, the case for Choice Two remains unaltered:
no matter how I choose now, there either is or is not a million pounds in
Box One, and either way I must be better off taking both boxes.

It is precisely in order to avoid this irrelevant complication—that my
present choice might affect what is in Box One, in that what is there
might vary with my choice—that some have added the proviso that the
contents of the boxes might be visible to others, though not to the Choo-
ser himself. They can see what is in the boxes, they can see that it does
not change, but, not knowing what they do, the problem for the Chooser

remains the same. Schlesinger, however, has used this version of the problem to generate a further argument for Choice Two: I can know that an intelligent, well-informed and perfect well-wisher, knowing what is in the boxes, would advise me to make Choice Two, both if Box One is full and if it is empty; what an intelligent, well-informed and perfect well-wisher would advise me to do must be in my best interests; so, whatever is in Box One, it must be in my best interests to make Choice Two.

What, then, is the case for Choice One? Schlesinger offers the following argument: In principle there seem to be the four following outcomes, in descending order of desirability:

A Predictor predicts Choice One, Chooser makes Choice Two, and gets a million plus a thousand;

B Predictor predicts Choice One, Chooser makes Choice One, and gets a million;

C Predictor predicts Choice Two, Chooser makes Choice Two, and gets a thousand;

D Predictor predicts Choice Two, Chooser makes Choice One, and gets nothing.

But in fact, given that the predictor is infallible, the first and last of these are not possible; and since Outcome B is clearly the better of the two that remain, Choice One is the choice to make. Notice, however, that this argument requires not merely, as Schlesinger puts it, that 'the Predictor . . . be believed virtually infallible' (p. 210), but that he actually be absolutely infallible, in the sense that there is no possibility of his ever going wrong. So long as there is that possibility, however slight, the Predictor is not absolutely infallible, Outcomes A and D cannot be ruled out *a priori*, and Schlesinger's argument for Choice One fails. And while the Predictor's past success and the principles of inductive logic certainly imply that the Predictor will be right on this occasion as on others, they do not guarantee that he is absolutely infallible in the required sense. The fact that he has never yet gone wrong is not by itself proof that error is impossible for him.

Nevertheless a slight modification might seem to restore Schlesinger's argument. In deciding how to choose the Chooser has, in effect, to decide between three alternative explanations of the Predictor's amazing performance:

I that there is some connection between the Predictor's predictions and the choosers' choices which accounts for his past success, and which will ensure his success on other occasions, such as this one;

II that there is some connection between the Predictor's predictions and the choosers' choices which accounts for his past success, without being such as to ensure his success on other occasions, such as this one;

III that there is no connection between the Predictor's predictions and the choosers' choices, so that his past success is sheer chance, mere coincidence, and hence does not ensure, or even suggest, his success on other occasions, such as this one.

More briefly: is the Predictor absolutely infallible (Explanation I), extraordinarily reliable (Explanation II), or incredibly lucky (Explanation III)?

Now if Explanation III were correct there would be no question that Choice Two is the right choice; the Chooser can choose in the certain knowledge that he cannot suffer from taking both boxes, in the hope that luck has finally run out for the Predictor, and with a 50% chance that Box One will be found to contain a million pounds. But unhappily the evidence is very much against Explanation III. Certainly luck is always a possibility, but the longer a run continues the more likely it is that it is not just luck, that there must be some explanation. If I roll 100 consecutive sixes with an unfamiliar dice it would be foolish not to anticipate another six on the next roll. Of course if I know the dice to be fair then I know that 100 consecutive sixes do nothing to increase the likelihood that the next roll will produce another six; but the 100 consecutive sixes make it more and more probable that the dice is loaded, that the result will be my 101st six. Similarly here: the principles of inductive logic, such as they are, make Explanation III the least likely explanation, and it would be foolish to make Choice Two on that ground, as foolish as betting against a six—or for that matter, betting at all— on a dice that has rolled 100 consecutive sixes. Whatever accounts for what is going on it is unlikely, though not impossible, that it is sheer chance.

But although the inductive evidence clearly favours Explanations I and II it does not so clearly discriminate between them. Schlesinger's argument for Choice One appeals, in effect, to Explanation I, but Explanation II remains a possibility. Nevertheless, if Explanation II does not, like Explanation I, rule out Outcomes A and D as impossible, it does render them unlikely. With more information about how the Predictor makes his predictions, and hence more knowledge of what might, on occasion, result in his going wrong, I would be in a better position to estimate the likelihood of his going wrong on this particular occasion. But since I lack this information I can only conclude that since he has never gone wrong before it is unlikely that he has gone wrong this time, although that does remain a possibility. So it does seem that in his present state of knowledge the Chooser will, by a modification of

Schlesinger's argument—it is, in effect, an informal version of that of Bar Hillel and Margalit—be better advised to accept that the Predictor is at least reliable, and so make Choice One.

Thus the inductive evidence seems to point in the direction of Choice One. But if induction favours Choice One, simple arithmetic and the advice of an intelligent and well informed perfect well wisher favour Choice Two. From which Schlesinger draws the disconcerting conclusion that induction breaks down here, that the essential unpredictability of free choice means that, contrary to all appearances, Explanation III must be correct after all.

III. A Solution

There is, however, a further error in Schlesinger's argument for Choice One, an error which persists even in its revised form. For if the Predictor really is absolutely infallible then it is not even the case, when the Chooser comes to make his choice, that both Outcome B and Outcome C remain open. Certainly Outcomes A and D can be ruled out *a priori*, given Explanation I, but once the Predictor has made his prediction then either Outcome B or Outcome C will also be impossible, though the Chooser does not know which. For once the Predictor has made his prediction, that prediction becomes fixed and unalterable: having made the one prediction, it is no longer possible for him to make the other. So given that the Predictor is absolutely infallible, it is at the time of choosing equally impossible, and in just the same sense, for the Chooser to make any choice other than that predicted. One alternative is now closed: the Chooser cannot choose in any other way. That is, if the Predictor is absolutely infallible in the required sense it is impossible for the Chooser to choose other than as predicted; once the prediction has been made it is impossible for there to be a different prediction; hence at the time of choosing only one Outcome is possible, the one where the Chooser makes that particular choice, whichever it is, which the Predictor has predicted.

This argument depends, of course, on the assumption that it is impossible to alter the past, to make it other than it has been. But it does not conflict with a more moderate conception of reverse causation, in which the past is held merely to depend on, or be determined by, the present or future. For even if the connection between prediction and choice runs, so to speak, from the latter to the former—and this may very well be the best explanation of the Predictor's infallibility—it will still be impossible for the choice to be other than that predicted, once it is impossible for the prediction itself to vary. If the choice determines the prediction, in that the prediction is as it is only because the choice is as it is, then once the prediction is fixed and unalterable, so too is the choice.

Thus even given Explanation I, Schlesinger's argument for Choice One fails. For once the Predictor has made his prediction his infallibility, the impossibility of his going wrong, ensures that the predicted choice is the only choice available, and *a fortiori* the best choice available. Moreover we can now see the error underlying the more general argument for Choice One. That argument might be put: if you make Choice One you will find a million pounds in Box One; if you make Choice Two you will find nothing in Box One; therefore you should make Choice One. But the premisses are ambiguous between 'If you can make Choice One, if that choice is possible for you, you will find Box One full, whereas if you can make Choice Two, if that choice is possible for you, you will find it empty', and 'If you were to make Choice One, *per impossibile* as it may be, you would find Box One full, whereas if you were to make Choice Two, *per impossibile* as it may be, you would find it empty'. Only on the second interpretation are the premisses a reason for preferring Choice One to Choice Two, but only on the first interpretation are they true, given Explanation I.

Under Explanation II the situation will be slightly different, for in this case the Predictor is not absolutely infallible and there will always be the possibility, however slight, that his prediction might on this occasion be falsified. To take one possibility, it may be that the Predictor has access to information from which the future choice can be deduced with complete accuracy, but for the prediction to be correct the Predictor has to do his calculations correctly: a mistake on his part could always result in a false prediction, even though no such error has occurred to date. Or to take another possibility, it may be that the factors on which the prediction is based are factors which make it extremely likely, though not absolutely certain, that the Chooser will make one choice rather than the other: here too there is always a chance that the Chooser will fail to choose as predicted. Notice that the second case suggests, as the first does not, that the Chooser does have some freedom of choice. In the first case his choice is determined, only one choice is available to him, every bit as much as under Explanation I, whereas in the second how the Chooser will choose does remain open, at least in small part, until he actually makes his choice. In either case it is possible that the Chooser will choose other than as predicted, but only in the second is it possible for the Chooser to choose in either way. But this difference does not affect the argument that Choice Two is the choice to make if you can; it only affects the question of whether you can. For if there is some small possibility of choosing other than as predicted, perhaps even some small scope for freedom of choice, then to that extent all four outcomes are possible, and inasmuch as Outcome A is preferable to B, and Outcome C to Outcome D, Choice Two is the choice to make, if you can.

Nevertheless the inductive evidence is that it is extremely unlikely

that, this time for the first time, the Chooser will falsify the prediction. But this is no argument for preferring Choice One. It is, rather, evidence that there is no choice about it, or that any scope for freedom of choice is so small as never yet to have been exercised. The existence of a Newcomb Predictor would be evidence that such choices are determined, or at any rate that our freedom of choice is extremely limited. Not, of course, that this provides any reason to think that our choices actually are determined; the very unlikelihood of a Newcomb Predictor suggests rather the opposite. But we can now see that the original dilemma was misstated. The fact that I have every reason to think that the Predictor will have predicted correctly on this occasion as on others does not give me any reason to make Choice One. Rather it gives me every reason to think I have no choice in the matter at all, or that if I do have any freedom it is a freedom I am unlikely to exercise.

Thus Choice Two is always the right choice, the choice to make if you can. As a rational man I would myself attempt Choice Two, confident that whether Box One is full or empty I cannot suffer from that choice, hopeful that I might at last demonstrate freedom of choice in the presence of a Newcomb Predictor, but resigned to receiving a mere thousand where less rational mortals, who from the evidence of *Scientific American* (March 1974, p. 102) outnumber the rational by the order of 5 to 2, stand to gain a million. Perhaps it is precisely because I am a rational man that the Predictor is able to predict my choice, and leave Box One empty. Perhaps, in this respect at least, I would be better off were I less rational than I am. The penalties of philosophy are no less than its pleasures.[1]

University of Warwick   © Don Locke 1978

CARNAP ON FREGE ON INDIRECT REFERENCE

By ALAN HOLLAND

IN § 30 of *Meaning and Necessity* (2nd edn., Chicago, 1956) Carnap draws attention to what he terms the 'disadvantages' of Frege's theory of names. Frege held that there belongs to a name both a sense, which it 'expresses', and a referent (or nominatum), which it designates. He also held that in indirect speech, as opposed to customary or direct speech, a name does not designate its customary referent but designates, instead, its customary sense. At the same time he attributed to a name occurring in indirect speech an indirect sense.

Carnap raises three objections:

(i) The theory leads to an infinite number of entities of 'new and unfamiliar kinds', and an infinite number of names for them.

(ii) The same name, in different contexts, may have an infinite number of different referents.

(iii) The same occurrence of a name may simultaneously have several different referents.

Although the second objection will be our chief concern, it will be convenient to begin with some remarks about the first, partly in order to illuminate the distinctness of these two objections.

Given a name '*a*' there is, standardly, an entity which it designates (the referent) and another entity which it expresses (the sense). If we wish to speak of this sense, we need a new name with which to designate it, for '*a*' designates the referent of '*a*', not its sense. Frege, too, noted this point and suggested a name, 'the sense of "*a*" ' (cf. 'On Sense and Reference', p. 59, in *Philosophical Writings of Gottlob Frege,* edd. Geach & Black, 2nd edn., Blackwell, 1970). However, according to Frege's theory there belongs to this new name, as to any name, a sense as well as a reference. And this sense must be a new entity distinct from the sense of '*a*', since the latter is already the referent of the name. To speak of this new sense we shall, again, need a new name, 'the sense of "the sense of '*a*' " ' to which there will, again, belong a sense; and so forth *ad infinitum*. As Carnap observes, this infinite number of senses is generated only on the principle that the reference and sense of a name must always differ. For otherwise there would be no warrant for inferring that the sense of 'the sense of "*a*" ' is an entity distinct from the sense of '*a*', and the proliferation of entities would thereby be halted. For the present let us assume the truth of the principle.

Even so the force of the objection seems weak. One could as well point to the proliferation of entities arising from the use of quotation

marks. For this device already provides the means of generating an infinite number of different names to which in turn belong, on Frege's theory, an infinite number of different senses. To speak of the name '*a*', to which some given sense belongs, we require a name ' "*a*" ' to which some further sense belongs; and so forth. The point is that the possibility of there being an infinite number of different senses, with an infinite number of different names for these senses, would not seem to constitute grounds for objection against Frege's theory *over and above* any grounds there may be for objecting to the notion of sense in the first place. Carnap speaks of customary sense as something 'familiar to us', yet goes on to speak of the infinite number of senses which is generated as comprising entities of 'new and unfamiliar kinds'. But the contrast is unwarranted. Each one of this infinite number of senses is an entity of a quite familiar kind, namely the customary sense of some name. And Carnap raises no objection to the notion of customary sense as such.

The threat to Frege's theory seems greater in the case of the second objection. Given Frege's claim that a name occurring in indirect speech designates the sense which it expresses in direct speech, the argument is that this makes available an infinite number of referents for the *same* name, answering to the availability of an infinite number of reiterations of the indirect speech context. Thus consider the sequence of sentences

(1) Scott is human

(2) It is possible that Scott is human

(3) John believes that it is possible that Scott is human

(4) It is not necessary that John believes that it is possible that Scott is human

which clearly can be continued indefinitely. According to Frege's theory the expression (name) 'Scott is human' in sentence (2) designates the sense of 'Scott is human' in sentence (1). It also expresses a sense. Assuming, as in the previous objection, that the sense and reference of a particular occurrence of a name must differ, the sense which it expresses must be a new sense distinct from the sense which it designates. This new sense will in turn be the referent of the expression 'Scott is human' in sentence (3), and so forth.

This argument assumes that a certain general principle governs Frege's approach to indirect contexts. Let us speak of customary or direct speech as being level zero of indirect speech. First level indirect speech is introduced by a single operator such as 'It is possible that . . .' or 'John believes that . . .'. Second, third and subsequent levels of indirect speech are introduced by applying the appropriate number of

operators, whether new ones or reiterations of ones already applied. Assuming an informal elucidation of 'sense' as 'method of identification', the principle in question is given thus:

P   The sense of an expression e in indirect context level n ($n>0$) is the method of identifying the sense of e in indirect context level $n - 1$.

For Carnap, the ground of objection is that Frege's method of semantical analysis 'results in a very complicated structure of the object-language' (p. 96). Dummett, however, derives from some remarks of Russell a much sharper ground of objection which effectively pre-empts that mentioned by Carnap (*Frege: Philosophy of Language*, London, 1973, p. 267). The point is simply that principle P yields no determinate entity, at any level, as the indirect sense of e. This is because the principle purports to elucidate the (indirect) sense of an expression solely in terms of its being *the* method of identifying a particular referent, in defiance of the manifest fact that for Frege there may be any number of different ways of identifying a particular referent. It is true that we know the sign which is to express this sense, namely 'e' itself. But, at the first level of indirect sense, the familiar customary sense, understanding of which initially gives the sign its significance for us, is already in employment as the referent of the sign. The indirect sense is to be, it appears, a new entity. Thus our knowledge and understanding of the sign which is to express this sense is of no avail. And if the first level of indirect sense is indeterminate, so too are all subsequent levels. It is true, also, that we know of other expressions, such as 'the customary sense of "e" ', which designate the very same referent as the sign whose sense puzzles us, and that we know what the senses of these other signs are. But we can no more identify the unknown sense with any of these other senses on the grounds that they constitute methods of identifying the same referent, than we can identify the senses of 'the evening star' and 'the morning star' with each other. (It should be noticed that the difficulty does not arise along with the first objection. For there, we know the new sign, e.g. 'the sense of "a" ', and furthermore we know its sense. Its sense is customary, and is a function of the customary senses of ' "a" ' and 'the sense of . . .').

Dummett's own response to this difficulty, a response anticipated but not developed by Carnap (op. cit., p. 129), is to jettison the notion of indirect sense entirely. Thus he writes (p. 268)

> . . . there is no such thing as the indirect sense of a word: there is just its sense, which determines it to have in transparent contexts a reference distinct from this sense, and in opaque contexts a referent which coincides with its sense.

Not only does this move dispel the difficulty, Carnap's version as well as

Dummett's own, but it is also, Dummett claims, in complete harmony with the rest of Frege's theory.

Two features of this emended theory require comment.

(1) It is an implication of the emended theory that one and the same sense can determine two different referents. Far from seeing this consequence as paradoxical, Dummett suggests that it was the misguided denial of its truth which led Frege to the notion of indirect sense in the first place. It only seems paradoxical, he argues, if one supposes that sense, all by itself and in isolation, determines reference. The truth is that, for Frege, sense determines reference only in the context of some sentence, so that reference is 'determined jointly by the sense of the word and the kind of context in which it occurs' (p. 268). Thus one and the same sense might indeed help to determine two different referents, if the two kinds of context involved are sufficiently different.

(2) It is an implication of the emended theory that in indirect speech the sense of an expression is identical with its referent. Oddly, Dummett seems to find this consequence unremarkable, even though Carnap, for one, thought that Frege was bound to reject it (op.cit., p. 129). At an intuitive level the problem is simply this. How can a method of identification be identical with what that method identifies? For an answer one might perhaps look to a distinction between two styles of designation for thoughts, which can be termed the 'allusive' and the 'revealing' styles. Thus we might compare the allusive style involved in 'Caesar believed what the soothsayer said' with the revealing style of 'Caesar believed that he should beware the Ides of March'. In the former case, but not in the latter, we need to be supplied with further information if we are to know what it was that Caesar believed. In the latter, the thought appears *in propria persona* (as Frege might have said): the words designate this thought precisely by means of expressing it. So far the answer is, perhaps, plausible. However, it does overlook the fact that, for Frege, not only are there different ways of designating the same referent; there are also different ways of expressing the same thought. This is a possibility which is not captured on the view that in indirect contexts the thought expressed and the thought designated are one and the same. 'The same sense has different expressions in different languages or even in the same language' writes Frege in 'On Sense and Reference' (op. cit., p. 58). And there it might seem as if trivial linguistic differences were in question such as that effected by the substitution of 'II' for '2'. But a passage in 'On Concept and Object' suggests otherwise. 'Language', says Frege, 'has means of presenting now one, now another, part of the thought as the subject' (op. cit., p. 49); and he instances the distinction of active and passive forms. What he seems to have in mind is that a given thought may be composed in two (or more) ways by virtue of some 'principle of compensation'. Each different expression of the same thought will

contain a part whose sense differs from the sense of any part of another expression of that thought. But identity of thought will be restored by compensatory differences of sense elsewhere. An example of two such expressions of a thought might be '2 < 3' and '3 > 2'. When Frege reverts to the point in 'The Thought', besides again mentioning the active-passive distinction he cites the class of token-reflexive sentences. Here, different circumstances of utterance, which contribute to determining the thought, compensate for the differences in sense of 'I' and 'you', 'today' and 'yesterday', and so forth. The existence of different ways of expressing the same thought, then, seems to be a significant semantic fact which ought ideally to be registered in a Fregean theory.

Thomas Baldwin, in ANALYSIS 35.3, brings to bear a more serious looking objection. He argues that Dummett's version of the Fregean theory makes genuine mistakes about the identity of propositions (thoughts) and concepts impossible. For, given the truth of the premises

X knows that the thought that p = the thought that p
The thought that p = the thought that q

Dummett is unable to block the inference to

X knows that the thought that p = the thought that q

The reason apparently has to do with the fact that identity of sense (i.e. in this case identity of thought) is always sufficient on the Frege-Dummett theory to license substitution in indirect speech.

It is not at all clear, though, how this objection is meant to work. The truth of

The thought that p = the thought that q

certainly licenses the substitution of any expression designating the thought that q for any expression designating the thought that p. But in the sentence

X knows that the thought that p = the thought that p

the expression 'the thought that p' does not designate the thought that p (in either of its occurrences). Instead, it designates the customary sense of the expression 'the thought that p'. Hence, substitution of 'the thought that q' is not licensed unless it can be shown that the customary sense of 'the thought that q' is identical with the customary sense of 'the thought that p'. Now, according to Dummett, in a context such as 'John believes that p' the sense of 'p' is identical with its referent (which is of course the thought that p), and therefore also identical with the sense of 'q' in 'John believes that q', if the thought that p = the thought that q. But it cannot be inferred that the customary sense of 'the thought that q' is identical with the customary sense of 'the thought that p'. For there is as yet no

warrant for assuming that the sense of 'p' (or 'q') in 'John believes that p (or q)' is the same as the customary sense of 'the thought that p (or q)'. Certainly the fact that the expressions concerned designate the same referent is not relevant for the familiar reason that reference does not determine sense. The point can be better appreciated if we adopt the different, 'allusive' style of designation. Given the truth of the premisses

> X knows that the last thought which occurred to Cromwell = the last thought which occurred to Cromwell

> The last thought which occurred to Cromwell = the last thought which occurred to Napoleon

nothing that Dummett says licenses the conclusion

> X knows that the last thought which occurred to Cromwell = the last thought which occurred to Napoleon.

Nevertheless, returning once more to the revealing style of designation, there are indeed conflicting pressures. On the one hand, if the thought that p = the thought that q and if therefore, as Dummett must hold, 'p' and 'q' in the context 'John believes that . . .' are identical in sense, whilst it may not follow directly that the customary sense of 'the thought that p' = the customary sense of 'the thought that q', it is difficult to see any grounds for supposing that they differ. Consider the comparable pair of expressions 'John's belief that p' and 'John's belief that q'. Given that 'p' and 'q' have the same sense as each other when preceded by 'John believes that . . .', it seems reasonable to hold that they also have the same sense as each other when preceded by 'John's belief that . . .'. Bearing in mind that the sense of an expression is a function of the sense of its parts, the conclusion is inevitable that 'John's belief that p' and 'John's belief that q' are identical in sense, since the remaining parts of the expressions coincide. The analogy between the forms of expression 'John's belief that p' and 'the thought that p' makes it almost equally inevitable that the customary sense of 'the thought that p' = the customary sense of 'the thought that q'. Hence, the objection that Dummett's theory does away with mistakes about propositional identity, at least when the propositions are designated in the revealing style, appears justified.

On the other hand it is precisely with respect to thoughts identified in the revealing style that one is inclined to wonder whether mistakes of identity are in fact possible. For the style of reference involved in the paradigms of identity mistakes is more akin to the style we have termed allusive. We can mistake a horse for a cow when one or other is 'allusively' identified as an animal glimpsed in the far corner of a field. We

cannot, or so Plato held (*Theaetetus* 190C), simply think that a horse is a cow, when each is thus 'revealingly' identified.

But perhaps enough has been said to make it worth raising the question whether there is any alternative to Dummett's theory. I think there is such an alternative and label it the 'literal' theory, on the grounds that it keeps close to what Frege actually says. (This is an explanation of the label, not an argument in favour of the theory). Frege *says* there is such a thing as indirect sense, and is naturally taken to be saying that it differs from customary sense. On the other hand he makes no mention of second or third levels of indirect sense such as Carnap says he is committed to. Now in order to render the notion of indirect sense viable we must face the problem of how it is to be determined other than through its referent. This is not so difficult, I suggest, if we identify it with the manner of expressing a thought discussed by Frege, as we saw, in 'On Concept and Object'. It needs to be noticed that we are not entirely in the dark about this problematic sense. We do not merely know the referent which it allegedly determines; we know also the sign which is to express this sense. Furthermore we understand the sign. This overall understanding of the sign, I claim, may readily be taken to *include* a grasp of the indirect sense, if by 'indirect sense' we mean the particular manner of expressing the thought which the sign embodies. It is a familiar Fregean point, after all, that customary sense does not exhaust the overall meaning which any expression has. In understanding a sign, then, we understand not only a certain method of identifying reality (the customary sense), but also a certain method of identifying a thought (the indirect sense). Thus 'Scott is human' and 'Humanity belongs to Scott' express the same method of identifying the True but different methods of identifying the particular thought that Scott is human.

This theory re-introduces the possibility of mistakes about propositional identity, even when the propositions in question are referred to in the revealing style. For, recalling our argument on this point, even though the thought that p =the thought that q, the senses of 'p' and 'q' in the context 'John believes that . . .' will be different. Hence the senses of 'John's belief that p' and 'John's belief that q' will be different, and so also the senses of 'the thought that p' and 'the thought that q'. Hence a person may be mistaken as to the identity of 'p' and 'q'. Furthermore, the theory suggests what the source of such a mistake would be. But it would be wrong to infer that the theory actually commits one to the view that such mistakes are possible. For it might be argued that not *any* kind of difference of sense is such as to make mistakes of identity possible.

It won't have passed unnoticed that if we prefer the literal theory to Dummett's theory we forfeit the advantage of his reply to Carnap's original objection. But another reply is available, namely that principle P, on which the objection is based, is a grossly unwarranted extrapolation

from Frege's procedures. None of the considerations which led Frege to introduce the notions of indirect sense and reference constitutes an argument for introducing further levels of indirectness. As Dummett himself observes, if substitution on the basis of identity of sense is adequate at the first level of indirect speech it is adequate too at any subsequent level. Again, the restriction to but one level of indirect sense and reference is quite adequate to meet the point that the truth of a whole sentence involving indirect speech is indifferent to the truth or falsity of the subordinate clause ('On Sense and Reference', p. 66). For that point simply requires the reference of a subordinate clause to be something other than a truth value. A more technical consideration is this. Suppose we view '( )^h' and '( )^p' as functions, to be read, roughly, as '. . . is human' and 'It is possible that . . .', yielding truth values for appropriate arguments. Let 'B_j . . .' be short for 'John believes that . . .'. By itself, then, '(Scott)^h' designates a truth value. In the sentence 'B_j(Scott)^h', it designates a thought. The change of designation here from truth-value to thought carries with it a change of designation of the part 'Scott'. Analogously, '(Scott is human)^p', designates a truth-value. In the sentence 'B_j(Scott is human)^p', it designates a thought. Now *should* we expect the change of designation here from truth value to thought to carry with *it* a change of designation of the part 'Scott is human'? It seems to me that we should not. We should expect this only if '( )^h' functions with 'Scott' in the same logical manner as '( )^p' functions with 'Scott is human'. But we know, precisely, that it does not. Thus not only is there no positive reason to postulate further levels of indirectness in Fregean theory. There is positive reason not to.

There are two ways of viewing the logical role of operators which introduce indirect speech when these are themselves embedded in an indirect context, which appear compatible with the rest of Frege's views. One is to suppose that they serve to seal off the thought which they introduce from the effect of the application of any further operator. The other is to suppose that they only effect the transformation of reference from truth-value to thought when they themselves are to the fore, and that once embedded they are powerless to effect the transformation. (Like bees, as it were, they sting only once.)

One final point. It is possible that one's preference for the literal theory or for Dummett's theory simply turns on what are supposed to be the conditions for the identity of thoughts. Dummett assumes relatively tight conditions akin to those embodied in Carnap's notion of intensional isomorphism (op. cit., p. 279). I have assumed more liberal conditions, allowing for transformations from active to passive and the like. The connection is simply illustrated by the reflection that if intensional isomorphism is required for identity of thought one is unlikely to be much interested in preserving the possibility of identity

mistakes with respect to thoughts. But interest may revive as the scope
for transformation is increased.

University of Lancaster   © ALAN HOLLAND 1978

ANIMAL RIGHTS: A REPLY TO FREY

By DALE JAMIESON *and* TOM REGAN

IN his paper 'Animal Rights' (ANALYSIS 37.4) R. G. Frey disputes
what he refers to as 'the most important . . . argument' for the view
that 'animals do have rights' (p. 186). Frey formulates the argument in
the following way:

(1) Each and every criterion for the possession of rights that excludes
    animals from the class of right-holders also excludes babies and the
    severely mentally-enfeebled from the class of right-holders;

(2) Babies and the severely mentally-enfeebled, however, do have
    rights and so fall within the class of right-holders;

(3) Therefore, each and every one of these animal-excluding criteria
    must be rejected as a criterion for the possession of rights
    (pp. 186-7).

Frey's position is that premiss (2) 'is not obvious and requires defence',
and that 'the best defences of it, *if they stand at all,* specifically exclude
animals from the class of right-holders' (p. 187). The dilemma this is
supposed to pose for those who argue for animal rights is this: 'either
premiss (2) cannot be defended or else premiss (1) . . . is false' (ib.).
'In either case,' Frey concludes, 'this most important argument in behalf
of animal rights fails' (ib.).

At the outset a serious question must arise as to whose argument, if
anyone's, Frey is attacking, when he disputes this 'important argument'.
'Instances of this argument abound,' he claims (p. 187), yet he fails to
document a single instance of it. The only supposed occurrence Frey
cites is from Andrew Linzey's *Animal Rights,* the allegedly incriminating
passage being the following:[1]

[1] Andrew Linzey, *Animal Rights* (London, 1976), p. 24. Frey fails to show that this
argument occurs in any of the other sources he cites. These sources are the following:
J. Feinberg, 'The Rights of Animals and Future Generations', in W. Blackstone (ed.),
Philosophy and Environmental Crisis (Athens, Georgia, 1974); S. & R. Godlovitch, J. Harris
(eds.), *Animals, Men and Morals* (London, 1971); T. Regan, 'The Moral Basis of Vegetarian-
ism', *Canadian Journal of Philosophy,* vol. V, 1975, pp. 181–214; T. Regan, P. Singer (eds.),
Animal Rights and Human Obligations, (Englewood Cliffs, New Jersey, 1976); P. Singer,
'Animal Liberation', *The New York Review of Books,* vol. XX, no. 5, April 5, 1973, pp. 17–21;
P. Singer, *Animal Liberation* (London, 1976).

If we accord moral rights on the basis of rationality, what of the status of newly born children, "low grade" mental patients, "intellectual cabbages", and so on? Logically, accepting this criterion, they must have no, or diminished, moral rights.

Questions about the soundness of Linzey's argument are one thing; those concerning its form are another. Frey would have us suppose that Linzey's argument provides an example of the form of argument Frey attacks. The passage just quoted from Linzey, however, provides us with no grounds for thinking this. Linzey argues that, given a particular criterion for right-possession (namely, rationality), not only animals, but also some humans will fail to qualify as possessors of rights, or will at most have 'diminished moral rights'; in the passage just quoted he does not allege, and neither does he commit himself to, the view set forth in Frey's premiss (1)—namely, '*Each and every* criterion for the possession of rights that excludes animals from the class of right-holders also excludes babies and the severely mentally-enfeebled from the class of right-holders.' And since it is this view which an exponent of what Frey calls 'the most important argument' for animal rights must affirm or be committed to, Frey fails to show even that Linzey argues in the way he (Frey) criticizes, let alone that 'instances of this argument abound.' Even were we to grant, therefore, that anyone who argues in the way Frey disputes must face the dilemma he sets out, Frey provides us with no evidence for supposing that any "animal rightist" argues in this way.[1]

Whether or not any animal rightist does argue after the fashion Frey criticizes, it is reasonably clear that none needs to or should. For there are criteria for the possession of rights which, if accepted, could permit us to ascribe rights to all humans and withhold them from all animals. If, for example, we suppose that possession of an immortal soul is necessary and sufficient for inclusion within the class of right-holders, *and* that all and only human beings have immortal souls, then we have a criterion which would exclude all animals but no human beings. It would be prudent, therefore, for animal-rightists not to subscribe to so unqualified a position as the one which Frey, in his premiss (1), attributes to them. All that an animal rightist should and need maintain, and all that, for example, Linzey, in the above quote, commits himself to, is the view that some humans will be excluded from the class of right-holders *given certain* criteria for the possession of rights that exclude animals,[2] not that

[1] Though not an argument, it may not be irrelevant to remark that the present authors do not know of a single instance of an animal rightist's arguing in the way Frey criticizes.
[2] This form of argument has recently come under attack. See Jan Narveson, 'Critical Notice' of Tom Regan and Peter Singer (eds.), *Animal Rights and Human Obligations* (Englewood Cliffs: Prentice-Hall, Incorporated) 1976; and Peter Singer, *Animal Liberation* (New York: A New York Review Book) 1975; in *The Canadian Journal of Philosophy*, July 1977. For a critical discussion of Narveson's objections to this form or argument, see Tom Regan, 'Narveson on Egoism and the Rights of Animals' in this same issue of *The Canadian Journal of Philosophy*.

some humans will be excluded given each and every criterion that ex-
cludes animals. What needs to be made clear, and what Frey evidently
fails to understand, is this: the fact that some humans will be excluded,
given certain criteria for right-possession, is *just one* reason (albeit a
vitally important one) which animal rightists can and do give for reject-
ing *some* criteria proposed for right-possession. It is not essential to their
position that this be the *only* reason they can give for rejecting *all* the
criteria they do or should reject, or that, though there may be other
reasons they *may* give, this is one they *must* give against every criterion
they do or should reject. If, again, we suppose that all and only human
beings have an immortal soul, then no animal rightist could object to
the view that right-possession is to be determined by soul-possession
because this criterion for right-possession excludes some humans from
the class of right-holders. But it does not follow from this that an animal
rightist cannot object to this criterion on some *other* grounds—for
example, that it is false that humans have immortal souls, or that there is
no plausible connection between having souls and having rights.

There exists, then, no apparent reason why any animal rightist must
or should accept Frey's premiss (1); and this, coupled with the fact that
Frey fails to demonstrate that any animal rightist does accept it, would
seem to constitute sufficient grounds for concluding that, so far is he
from identifying and criticizing 'the most important argument' for ani-
mal rights, the argument Frey actually examines is of no importance
whatsoever for the case for animal rights.

Although Frey appears to be mistaken in supposing that premiss (1)
is crucial to any argument for animal rights, it does not follow that his
principal thesis (hereafter referred to as 'Frey's thesis') is mistaken—the
thesis, namely, that 'the best defences of [premiss (2)], *if they stand at all*,
specifically exclude animals from the class of right-holders.' Although it
does not follow that Frey's thesis is false, there are other grounds for
thinking that it is, or at least that Frey himself fails to show that it is
true.

A "defence" of premiss (2), as Frey uses the concept, is an attempt to
show that infants and the severely mentally-enfeebled satisfy some re-
quirement (e.g., rationality) proposed as a criterion for the possession
of rights. 'The best defences' of this premiss, according to Frey, are (i)
'the potentiality argument', wherein one argues that, though not actually
in possession of some capacity allegedly required for right possession
(say, rationality), infants have this capacity potentially and so should be
accorded rights; (ii) 'the similarity argument', wherein it is argued that,
though lacking the allegedly required capacity both actually and poten-
tially, the severely mentally-enfeebled should be accorded rights be-
cause they 'betray strong similarities to other members of our species';
and (iii) a religious argument, according to which rights are to be

attributed only to beings possessing an immortal soul, a blessing which all humans may be alleged to enjoy.

Frey's thesis, then, appears to consist of two parts: (a) there is his view that the potentiality, similarity and religious arguments are, in his words, 'the best defences' of premiss (2); and (b) there is his view that these defences, 'if they stand at all, specifically exclude animals from the class of right-holders'. Now, this latter claim is ambiguous, as is Frey's use of 'animals' throughout his essay. It could be interpreted to read either that these defences, if they stand at all, exclude some *or* all animals from this class. Since Frey nowhere concedes that these allegedly 'best defences' might be sufficient rationally to ground the view that at least some non-human animals have rights, it does not seem contrary to the spirit of his essay to interpret part (b) of his thesis to read that each of these defences, if it stands at all, specifically excludes *all* animals from the class of right-holders.

Interpreted thus, part (b) is at least highly controversial. If we suppose that part of what it is to be rational[1] is to be able to make inferences; to be able to select the most efficient means of achieving certain ends; to be able to symbolize; to be able to analyse concepts into their constituent features; and to be able to recognize instances of general concepts; then it is at least arguable that some infant animals (e.g., infant primates) are, to some degree, 'potentially rational', in which case the potentiality argument, 'if it stands at all' as a defence of the view that (normal) human infants have rights, does not, or at least does not appear to, 'specifically exclude (all) animals from the class of right-holders'.

Similarly in the case of the similarity argument: given any basis (for example, physical appearance) in terms of which to judge the relative similarity of other beings to normal human beings, cases will arise where some animals will betray stronger similarities to these paradigm humans than do some non-paradigmatic humans (e.g., some physically deformed humans who are severely mentally-enfeebled). Thus, if these latter humans should be accorded rights, on the grounds of their similarity to these paradigm humans, then those animals who betray the same or a greater degree of similarity ought also to be accorded rights. *If*, then, the similarity argument 'stands at all', it does not, or at least it does not in any clear way, 'specifically exclude (all) animals from the class of right-holders'. The case of the religious argument may be somewhat different, but enough already has been said, without going into attempts to ground rights on the existence or non-existence of souls, to make it clear that it is at least arguable that some animals should be accorded rights, if the potentiality and similarity arguments happen, in Frey's words, to 'stand at all'. And this is enough to cast doubt on part (b) of his thesis.

[1] On this topic, see D. Premack, 'On Animal Intelligence', in H. Jerison (ed.), *Perspectives on Intelligence* (New York: Appleton-Century-Croft), forthcoming.

It remains to be asked whether part (a) of Frey's thesis is true: are 'the best defences' of premiss (2) the ones Frey considers? The issues here are complicated. They involve fundamental questions about how to evaluate proposed criteria for right-possession. Short of answering these questions, it is difficult to see how some one or some few defences can non-arbitrarily be singled out as "the best". If, for example, possession of an immortal soul simply is not a reasonable criterion on which to base possession of rights, then the fact, if it is a fact, that all human beings, including infants and the severely mentally-enfeebled, have souls could not be construed as one of 'the best defences' of premiss (2); it could not be construed as a defence at all. Judgments about which defences of this premiss are "the best", therefore, would seem to presuppose very careful supporting argument. In this, if in no other respect, Frey's essay is disappointing. It is not possible to find fault with his supporting argument because it is not possible to find one. Why just the potentiality, similarity and religious arguments are "the best", he does not say. Equally importantly, why the fact that infants and the severely mentally-enfeebled are *sentient* does not count as one of 'the best defences' of premiss (2) is left undiscussed by Frey, an inexplicable omission since (i) what Frey professes to be considering, in his words, are ' the best defences' of that premiss, and (ii) it is the capacity for sentience on which some animal rightists (and Linzey in particular!) rely[1] as 'the best defence' of the view that infants and the severely mentally-enfeebled do have rights. It is of more than passing interest, moreover, that this same capacity (that is, sentience) is relied upon by these thinkers for including animals within the class of right-holders.

We conclude, therefore, that Frey fails to show that anyone who argues for animal rights has argued, or should or must argue, after the fashion he examines, *and* that he provides us with insufficient reasons, and sometimes no reasons at all, for accepting either part of his thesis. Arguments for animal rights may fail, but Frey fails to show that they do or, if they do, why they do.

North Carolina State University at Raleigh
© DALE JAMIESON *and* TOM REGAN 1978

[1] See, for example, Linzey, op. cit., p. 20 ff., and Singer, *Animal Liberation*, op. cit., p. 1 ff.

STATISTICAL JUSTIFICATIONS OF DISCRIMINATION

By Robert L. Simon

(a) 80% of the candidates who score poorly on Test T will fail to graduate from the university due to academic deficiencies. Therefore, the university is justified in refusing to admit *any* candidate who scores poorly on T.

(b) 80% of women lack the physical strength to perform job J. Therefore, employers are justified in refusing to consider *any* women as candidates for J.

(c) 80% of the members of group G are victims of past injustice. Therefore, compensatory programs are justified in providing benefits to *all* members of G.

SARA Ann Ketchum and Christine Pierce recently criticized arguments of kind (b) (Analysis, 36.2, pp. 91–95). They allow that 'statistical differences between the sexes would indicate that justice does not require equal distribution of women and men . . . within given job categories'. But they deny that 'all women may justifiably be discriminated against in hiring if sufficiently fewer women than men are qualified' (pp. 92, 93). On the other hand, James Nickel has defended arguments of kind (c) (Analysis, 34.5, pp. 154–160). Nickel argues that while preferential discrimination in favour of black persons as a group may not be justifiable by ideal principles of compensatory justice, it can be justified by administrative or pragmatic factors. Where there is a high correlation between being a member of a particular group and being a victim of social injustice, there is a pragmatic justification for compensating all members of the group in question. It would just be too costly and difficult to evaluate cases on an individual basis. Nickel concedes that the pragmatic approach 'may result in a certain degree of unfairness' but maintains that 'it does help to decrease administrative costs so that more resources can be directed to those in need' (James Nickel, 'Classification by race in compensatory programs', *Ethics*, 84.2, 1974, p. 148; see also his 'Preferential policies in hiring and admissions', *Columbia Law Review*, 75, p. 534).

Can one justifiably reject arguments such as (b), which use statistical generalizations to justify discrimination *against* a particular group, yet accept arguments such as (c), which use similar generalizations to justify discrimination *in favour* of a particular group? I will argue that to the extent that one agrees with Ketchum and Pierce in rejecting the use of statistical generalizations to justify discrimination against persons of a particular sex or race, one is committed to rejection of Nickel's pragmatic

defence of preferential discrimination as well. Either there is no dif-
ference between the two or, if there are differences, they are not such as
to justify any difference in evaluation of the arguments.

Consider an at least relatively benign use of statistical generalizations,
such as that found in (a). Surely, the university's policy is an acceptable
one, at least if either it is impossible to tell whether any given candidate
of the kind in question will wind up in the 80% who fail or the 20% who
succeed, or if it is excessively costly to make such fine distinctions, and
no other relevant information about the candidates is easily obtainable.

If such pragmatic considerations are sufficient to justify the use of
statistical generalizations as a basis for discrimination in all contexts then
(b) and (c) would be as acceptable as (a). Discrimination for and against
particular groups would be justifiable on cost accounting grounds alone.
But then, the price of accepting Nickel's argument for preferential
discrimination, given the appropriate empirical claims about efficiency,
is that we are committed to accepting arguments of kind (b), given
similar empirical claims about efficiency. But if, as is surely the case,
arguments from efficiency do not justify discrimination against women or
blacks, then arguments from efficiency for preferential discrimination in
favour of women and blacks are unacceptable as well (at least where such
preferential treatment in favour of group members requires a denial
of positions to non-members).

One might reply, however, that there is an important difference
between (a) and (c), on the one hand, and (b) on the other. For it can be
argued that discrimination directed against persons of a particular race
or sex as such is inherently invidious and degrading and so cannot be
justified by arguments that may be acceptable in more innocuous
contexts. Although Nickel's pragmatic approach would disfavour some
white persons, its *intent* is not to discriminate against whites as such but
rather is to extend certain benefits to the unjustly victimized. Similarly,
the intent of the university in (a) is not to practise systematic and
pervasive discrimination against previously low achievers who could
nevertheless do university level work. Rather, it is to serve more ef-
ficiently the university's high priority educational goals.

This reply will not do, however, for (b) would remain objectionable
even if the intent of employers is not to discriminate against women as
such but only to maximize profits. To distinguish in the way suggested
between (a) and (c), on the one hand, and (b) on the other, is to counten-
ance discrimination against a particular race or sex whenever the purpose
of such discrimination is in itself benign, e.g. to maximize profits.

Nickel himself argues that correlations appealed to by racists (and
presumably by sexists as well) are spurious (ANALYSIS, 34.5, p. 156f).
However, Ketchum and Pierce acknowledge that there may in fact be
some statistical differences between the sexes in areas relevant to job

qualifications. And some tests used to select among job applicants have had an adversely disproportionate effect on blacks.[1] So the issue of whether statistical generalizations justify discrimination against individuals remains even when false and invidious generalizations cited by racists are left out of the picture.

Perhaps, however, the difference between (b) and (c) lies not in the *intent* to discriminate but in actual discriminatory effect. That is, given past invidious discrimination against a group, even benignly motivated discrimination now practised against that group is likely to stigmatize members of the affected group and, at the very least, continue to perpetuate that group's unfairly imposed subordinate position. Thus, discrimination against blacks is so objectionable precisely because it contributes to or at least perpetuates the effect of past discrimination in creating an especially disadvantaged caste. Discrimination against whites has no such effect, for whites do not constitute an especially disadvantaged caste, nor is preferential discrimination likely to reduce them to such a state.

While such an argument might be quite strong in a world of sharp and clear caste distinctions, contemporary western society is not such a world. For what the argument implies is 'that the white majority is monolithic and so politically powerful as not to require the . . . safeguards afforded minority racial groups. But the white majority is pluralistic, containing within itself a multitude of religious and ethnic minorities—Catholics, Jews, Italians, Irish, Poles, and many others who are vulnerable to prejudice and who to this day suffer from the effects of past discrimination'.[2] Discrimination against individual members of such groups may perpetuate the psychological, social and economic effects of past discrimination. And while these effects may be more serious on the average for blacks, it is just the propriety of jumping from premises about the average to conclusions about individuals that is at issue.

Finally, it may be suggested that the difference between (b) and (c) lies in the difference in cost of obtaining the data that would make reliance on statistical generalizations unnecessary. Thus, it is relatively easy to test individuals for strength but relatively difficult to find out just what degree of injustice a particular individual may have suffered. Thus, (b) is unacceptable precisely because it is so easy to avoid appeals

[1] Whether such tests are constitutional if there is no intent on the part of those giving it to discriminate is an issue that has been considered by the United States Supreme Court. In Washington *v.* Davis, the Court held that it was not sufficient to show simply that more black than white applicants had failed a qualifying examination for police recruits; in addition, the Court required that the plaintiffs demonstrate a racially discriminatory purpose.

[2] This passage, quoted from Professor Lavinsky's contribution to the 'DeFunis Symposium', *Columbia Law Review* (1975), 520, 527, was quoted favourably by the Supreme Court of California in *Allan Bakke v. the Regents of the State of California* in which it overthrew the University of California at Davis's programme of preferential admission to medical school for black applicants.

to statistical generalizations in the context at hand. On the other hand, (c) at least has a plausible claim to acceptability just because of the difficulty of getting the relevant information about individual cases.

While there may well be this difference between (b) and (c), it is not a significant difference. Thus, we would find arguments of kind (b) objectionable even if it was quite difficult to verify claims about the strength of particular individuals. This can be brought out by comparing (b) and (a). Surely, the savings in cost would have to be far higher to warrant appeal to statistical generalizations to justify discrimination against women or blacks than to justify the use of standardized tests in university admissions.

This can be brought out more clearly is we consider an alternative to the use of Test T in the admissions process. Suppose that the university find that the presence in an applicant's background of (i) a family which has been mired in poverty for generations, (ii) a father who lacks a university education and (iii) a family history of diabetes is an excellent predictor of academic failure. Suppose further that it actually is less costly to verify the presence of these factors in particular cases than to administer T. But even given the above suppositions, we surely would think it *unfair* if the university promulgated a rule to the effect that candidates who satisfied (i), (ii) and (iii) would be eliminated from consideration *without even being given the chance to take the test*. On the contrary, each individual at least has a claim to be considered on his or her individual abilities and talents. Even if (i), (ii) and (iii) turn out to be better predictors than T, T is morally relevant to the making of academic decisions in a way that they are not. It is the refusal to allow certain students even to try for admission, when based on ascriptive grounds, that is unjust or unfair. Thus, it is not simply the *cost* of obtaining information that is at stake. The *kind* of information obtained and employed is morally relevant as well. (Similarly, even if it could be shown that certain genetic irregularities were excellent indicators of criminal tendencies in individuals, detention of individuals with such genetic irregularities prior to the commission of any crime would be unfair. This is particularly true if the purpose of such prior detention was simply to cut the costs of police work involved in finding a fleeing criminal.)

Likewise, since (b) and (c) replace consideration of morally relevant data with identification of ascriptive characteristics, they are similar in a morally significant way. Each involves a significant degree of unfairness to individuals. Now, it may be that in some contexts avoidance of excessive costs should take precedence over avoidance of unfairness. But if that is so, avoidance of the same amount of cost will count as much in favour of discrimination against blacks and women as in favour of discrimination benefiting blacks and women.

Accordingly, even if (b) and (c) are different in the way suggested, it is far from clear that this difference is sufficient to demonstrate the moral acceptability of (c) and the moral unacceptability of (b). The gains in efficiency involved may not be sufficient to compensate for the unfairness to individuals. And, at the very least, the price of opting for efficiency with respect to (c) is that of acknowledging that similar gains in efficiency would warrant discrimination against the very groups compensatory policies are designed to benefit. Given the nature of the price, we should be especially wary of accepting the claims of efficiency at the price of fairness.

Isn't there a difference, however, between sacrificing justice to efficiency *in order to right a wrong* and making a similar sacrifice for efficiency's sake alone? And isn't *that* the difference between (b) and (c)? That is, Nickel's argument may be interpreted as having moral as well as pragmatic force. The point of using Nickel's approach, on this interpretation, is to minimize injustice by providing compensation to a high proportion of those who are entitled to it at costs which impose a less severe burden on the rest of the community than would implementation of ideal principles of compensatory justice.

However, this kind of response begs one of the crucial issues at stake. For surely one of the major points at issue in the debate over preferential discrimination is whether such treatment creates new injustice by substituting group considerations for evaluation of individual cases. Indeed if, as I have suggested, the substitution of ascriptive criteria for consideration of individual cases does involve a significant degree of unfairness to individuals, then it will not be easy to show that the pragmatic approach minimizes injustice. Nickel is not entitled simply to *assume* that any injustice involved is outweighed *on the scales of justice* by the compensatory benefits provided. That the pragmatic approach cuts costs of administration is indeed initially plausible. That it minimizes injustice as well is far more controversial and thus requires significant additional support.

The price, then of shifting from a strategy of increasing efficiency to one of minimizing injustice is that of decreased initial credibility and consequent assumption of a far heavier burden of proof.

Perhaps that burden can be met. Or perhaps alternate justifications of preferential treatment for groups are acceptable. It seems to me that those are the issues on which debate should focus. Appeal to statistical generalizations will not help us here. If such appeals are understood as appeals to efficiency, they are vulnerable to the sorts of objections raised earlier. If they are understood as appeals to justice, they simply raise at a new level the very questions about fairness to individuals that the appeal to the administrative approach was designed to help us avoid.

Logically, then, one cannot have it both ways. If purely pragmatic

appeal to statistical generalizations can justify preferential discrimination in favour of women and black persons, then, given corresponding gains in efficiency, it can justify discrimination against those groups as well. In each kind of case, fairness to individuals requires consideration of relevant factors, but in each kind of case the appeal is to ascriptive classifications instead. But the price of appealing pragmatically to ascriptive considerations in one kind of case is that of logical commitment to similar appeal in similar contexts in the other kind of case. Even given significant gains in predictive power and efficiency, this sort of commitment is one of which we ought to be exceedingly wary. The moral status of discrimination remains the same, even though its victims and beneficiaries may change.[1]

Hamilton College   © ROBERT L. SIMON 1977

AN INVALID EPISTEMOLOGICAL ARGUMENT AGAINST DOUBLE-ACTION THEORIES

By NICHOLAS GRIFFIN

IT'S certainly true that psychological 'double-action' theories—theories which seek to explain overt or physical actions in terms of simultaneous or slightly prior covert or mental actions—are very unfashionable today. Nonetheless, as I shall point out, some of the arguments characteristically brought against them involve fairly elementary logical mistakes.

Consider, for example, the following argument from Ryle:

> It is admitted that one person can never witness the volition of another; he can only infer from an observed overt action to the volition from which it resulted, and then only if he has any good reason to believe that the overt action was a voluntary action, and not a reflex or habitual action, or one resulting from some external cause. It follows that no judge, schoolmaster, or parent ever knows that the actions which he judges merit praise or blame; for he cannot do better than guess that the action was willed.[1]

[1] Gilbert Ryle, *The Concept of Mind*, Hutchinson 1949, pp. 65–6; Penguin, 1966, p. 64. Ryle uses the same argument again, in a different context, on p. 87. (It should be noted that I am not concerned here with Ryle's puzzling slide from 'infer' to 'guess' in the passage quoted. The slide serves rather to obscure the point I want to get at.)

Of course, judges and schoolmasters do know when to assign praise or blame so the conclusion to be inferred is that the double-action theory is wrong. Ryle is arguing as follows: If the double-action theory is correct

    (1) If John does A intentionally then John performs an act of volition.

But

    (2) We don't know that John performs an act of volition.

Therefore:

    (3) We don't know that John does A intentionally.

But (3) is false, while (2) is true. Therefore, (1) is false. But (1) is the central postulate of the double-action theory, so the double-action theory is false.

    Jonathan Bennett uses a similar argument in order to refute a theory of language which he attributes to Locke, in which the difference between a mere utterance and a meaningful utterance is to be found in a covert act of meaning which attends the latter. He writes:

> [We] do not in fact know what inner acts, if any, [other people] perform. So, if an inner-act theory of meaning is correct, we do not yet know that people sometimes do, and that other animals sometimes do not, mean something by what they utter, and *a fortiori* we do not yet know *what* anyone means by what he utters.[1]

Bennett's argument can be reconstructed as follows: If Locke's theory of language is correct

    (4) If Joan utters S meaningfully then Joan performs an act of meaning.

And since

    (5) We know that Joan utters S meaningfully,

it follows

    (6) We know that Joan performs an act of meaning.

But since (6) is false and (5) is true it follows that (4) is false and thus that Locke's theory of language is false. In what follows I shall consider only Bennett's argument—exactly similar objections could also be made against Ryle's.

    When spelled out as above the invalidity of the argument (4)–(6) is fairly obvious: it is not an instance of *modus ponens*, and arguments of the form

---

[1] Jonathan Bennett, *Locke, Berkeley, Hume: Central Themes* (Oxford: Oxford University Press, 1971), p. 5.

$p \supset q$

We know that p

Therefore: We know that q

(of which (4)–(6) is an instance) are notoriously invalid. For example, from

(7) If R is the rule for finding square roots then 25 is the square root of 625.

and

(8) We know that R is the rule for finding square roots.

it doesn't follow

(9) We know that 25 is the square root of 625.

since, plainly, we may not have performed the necessary calculations.

Bennett might try to repair his argument in two ways, but neither repair, as I'll show, will give him the conclusion he wants. Firstly, Bennett might replace (4)–(6) by a genuine instance of *modus ponens*, namely:

(4a) If Joan utters S meaningfully then Joan performs an act of meaning.

(5a) Joan utters S meaningfully.

Therefore:

(6a) Joan performs an act of meaning.

But Bennett can now only use (4a)–(6a) to refute Locke's theory of language if he can show that (6a) is false while (5a) is true. But this is precisely what his original argument was intended to show. So the first repair leads, not to an argument against Locke's theory, but to mere counter-assertion.

Bennett's second possible repair is to replace (4)–(6) by:

(4b) We know that if Joan utters S meaningfully then Joan performs an act of meaning.

(5b) We know that Joan utters S meaningfully.

Therefore

(6b) We know that Joan performs an act of meaning.

(4b)–(6b) is still not valid, unless we can use

(10) $\Box(p \supset q) \supset (\Box p \supset \Box q)$

(which is a thesis of many modal systems, including all the Lewis ones) to transform (4b) into

> (4b′) If we know that Joan utters S meaningfully then we know that Joan performs an act of meaning.

Although the argument (4b′)–(6b) is valid, it does not help Bennett in his attack on Locke, for two reasons. First, from the fact that (10) is a reasonable thesis to have in systems of alethic modalities, it does not follow that it is a reasonable thesis of systems of epistemic modalities.[1] There are good grounds for believing that (10) is, in fact, false when '$\Box$' is read as 'it is known by X that . . .' For example, in trying to prove a proposition 'q' a mathematician might find himself able to prove only that $p \supset q$, and many years later it may be that he discovers a proof for 'p' but nonetheless never comes to know that q since he never comes to connect his later proof with his earlier one. Moreover, allowance has to be made for such perverse thinkers as Lewis Carroll's tortoise,[2] who refuse to make use of *modus ponens*. If (10) is rejected for systems of epistemic logic then Bennett cannot obtain the argument (4b′)–(6b) which he needs.

However, even if we granted Bennett (10), his argument is still subject to a second difficulty. Given the valid argument (4b′)–(6b), Bennett can argue that since (6b) is false and (5b) true it follows that (4b′) is false. But from this conclusion it doesn't follow that Locke's theory of language is false, because (4b′) is not a consequence of that theory. Accordingly nothing Bennett has done has shown any defect in Locke's theory: neither his original argument nor its two possible replacements are sound.

Moreover, since Ryle's argument is exactly analogous to Bennett's it, too, will fail for exactly similar reasons. In general, no epistemological argument along these lines is effective against any form of double-action theory.

McMaster University   © Nicholas Griffin 1978

---

[1] It nonetheless appears in a number of them, e.g., in Pap's system, 'Belief and Propositions', *Philosophy of Science*, 24 (1957), pp. 123–36; Rescher's epistemic system o in *Studies in Modality* (Oxford: Blackwell, 1974), p. 104; and Snyder's system in *Modal Logic and its Applications* (New York: Van Nostrand Reinhold, 1971), p. 201.

[2] Lewis Carroll, 'What the Tortoise said to Achilles', *Mind*, 4 (1895), pp. 278–80.

A PROBLEM IN THE JUSTIFICATION OF DEMOCRACY

By J. L. GORMAN

PROFESSOR Anscombe has presented a voting table of which, she claims, the following description is true: the majority is frustrated by the majority's will being fulfilled (ANALYSIS 36.4, June, 1976). This alleged paradoxical (although not contradictory) fact may, she suggests, make nonsense of democracy. She examines a plausible justification for democracy, argues that this justification is damaged by the paradox, and concludes that a particular technique of tyranny is possible within such a system.

Miss Anscombe's argument is too concise, and is in need of considerable clarification. Mr Leahy's contentment with the fact that there are more satisfied desires than unsatisfied desires in Miss Anscombe's table shows that some crucial points remain obscure (ANALYSIS 37.2, January, 1977). I propose to make explicit some of these points, and to argue that a significant problem exists. This problem does not, however, permit the specific tyranny described by Miss Anscombe.

The argument will be clearer with a simpler table. The simplest possible table that will exemplify Miss Anscombe's point will involve the minimum number of voters, the minimum number of questions, and the minimum majorities, such that (I) each question is decided by simple majority vote, and (II) the majority of the voters vote in the minority on the majority of the questions. In Miss Anscombe's table, the minimum majority of voters, A to F, vote in the minority on more than the minimum majority of the questions. Her example is therefore stronger than it needs to be in order to exemplify her claim. The simplest possible example is as follows:

		Voters				
		A	B	C	D	E
Questions	(i)	1	1	1	−1	−1
	(ii)	1	1	−1	1	−1
	(iii)	1	1	−1	−1	1

Note that '1' and '−1' do not mean 'yes' and 'no', respectively, in answer to some question. '1' represents a vote which is in accordance with the result: a successful vote. Similarly, '−1' represents an unsuccessful vote. Requirement (I) implies that it is necessary that there are more successful than unsuccessful votes.

Requirement (I) reflects our ordinary unconsidered understanding of majority rule or democracy. When requirement (I) is met, we might commonly say that the majority has its will fulfilled.

For any one voter considered individually, however, it is slightly

less clear in what sense he may be said to have his will fulfilled. We may suppose that his will is fulfilled only if all his votes are successful. It follows that the individual voter is frustrated if just one of his votes is unsuccessful: if, therefore, he votes in the minority on the minority of the questions. But Miss Anscombe is not concerned with this situation, for she regards frustration as occurring only when the majority of the voter's votes are unsuccessful. It is too much to ask that all his votes should be successful.

We have to assume that a voter is satisfied if he has more successful than unsuccessful votes, and that he is frustrated if he receives fewer successful than unsuccessful votes. In assuming this, we presuppose that each successful vote is equivalent to each other successful vote in terms of desire satisfaction received, and that each unsuccessful vote represents an equal cancelling amount of dissatisfaction. This typical hedonist utilitarian assumption is required in order to interpret the table as Miss Anscombe wishes to do. We may therefore regard '1' as a unit of satisfaction, say as representing one satisfied desire. Similarly, ' − 1' represents one dissatisfied desire. By using ' − 1' instead of Miss Anscombe's 'o' we are able to sum over the desires in each column of the table in an arithmetical way. Now that a successful vote is interpreted as a satisfied desire, it follows from requirement (I) that it is necessary that there are more satisfied than unsatisfied desires in Miss Anscombe's table. Nevertheless, the majority of the people can be dissatisfied for the majority of the time.

Implicit in Miss Anscombe's paper is the claim that the majority's will is fulfilled (the second part of the title to her paper) just when each question is decided by simple majority vote—requirement (I).

But also implicit is the claim that the majority is frustrated (the first part of the title to her paper) just when the majority of the voters vote in the minority on the majority of the questions—requirement (II)[1].

These two claims may be taken as proposed definitions. If we accept them, and if we accept that frustration is the contradictory of will-fulfilment, and given (as we are) that requirements (I) and (II) are met, then the following paradoxical description is true: the majority is frustrated by the majority's will being fulfilled. A strict contradiction does not arise, however, as 'majority' refers to a different set of people in each of its two occurrences.

I shall assume, I believe plausibly, that frustration is the contradictory of will-fulfilment. I argue from the second definition that the majority has its will fulfilled just when the majority of the voters vote successfully on a majority of the questions. In the simple table provided,

---

[1] I take 'in the minority' here to mean 'unsuccessfully', but note that this equivalence is plausible only while requirement (I) continues to be met. I owe this note of clarification to discussion with Professor J. A. Faris.

this situation could occur by forcing an 'undemocratic' decision in answer to, say, question (i). The table would then necessarily read as follows:

		Voters				
		A	B	C	D	E
	(i)	— I	— I	— I	I	I
Questions	(ii)	I	I	— I	I	— I
	(iii)	I	I	— I	— I	I

If we understand 'democracy' to be represented by requirement (I), then this situation is not democratic. But we do have a situation in which the majority of the voters vote successfully on the majority of the questions.

This shows us that we have two possible and conflicting criteria for saying 'the majority has its will fulfilled'. (Again, the word 'majority' refers to a different set of people for each criterion and there is no formal contradiction.) One problem which arises is, which criterion best fits what we ordinarily recognize as democratic? There seems little doubt that requirement (I) fits our current sense of democracy most clearly. But far more important is the question of justification: should there be more successful votes than unsuccessful votes for each question? Or should there be more successful votes than unsuccessful votes for the majority of voters? This is the crucial issue raised by Miss Anscombe's paper.

The conflict is brought out far more clearly in a table of the following kind:

		Voters				
		A	B	C	D	E
	(i)	— I	— I	— I	I	I
	(ii)	I	I	— I	— I	— I
Questions	(iii)	I	I	— I	I	— I
	(iv)	I	I	— I	— I	I
	(v)	— I	— I	— I	I	I

In this table the satisfaction of voters conflicts with the satisfaction of desires. Four of the voters, A, B, D, & E, are satisfied: they each had successful votes on three out of five questions. But only two questions were decided by majority vote. In addition, there are more unsuccessful votes than successful votes. Therefore the majority of desires is unsatisfied, although the majority of voters is satisfied. We can re-phrase the above problem of justification: should there be more satisfied than unsatisfied desires? Or should there be more satisfied than unsatisfied voters?

Miss Anscombe has offered a utilitarian justification for democracy.

Value, we accept, lies in the satisfaction of people's desires. Best value then appears to lie with the satisfaction of the greatest possible number, that is, with the satisfaction of the majority. This procedure also involves fairness: no one person's desires are allowed to outweigh another's.

But do we attempt to satisfy the greatest possible number of desires or of people? The principle does not tell us: it is ambiguous between two senses of 'the satisfaction of people's desires, each being given equal weight'. It might mean that desire satisfaction should be given equally to each person, or it might mean that each desire should be given equal weight with each other desire. The table shows the incoherence of attempting to maximize both of these.

Majority decision involves the view that each desire should be given the same weight as each other desire. This amounts to no more than an intention to use units of satisfaction: to count desires. There is no element of fairness in the matter of the distribution of desire-satisfaction between persons. This is a very thin kind of justification for majority decision. Surely it would be more nearly ideal to give desire satisfaction to each person equally, but this strong justification supports only the view that there should be more successful than unsuccessful votes for the majority of voters. Quite simply, majority decision cannot be relied on in principle to give us a fair distribution of desire-satisfaction between people; and only the latter deserves to be called democracy.

Let us suppose that the only justifiable aim is to maximize the happiness of the majority of the voters. This involves ensuring that the majority of voters have more satisfied than unsatisfied desires. In order to do this, we need to take some number of questions over which the calculation is to be made. The liberal utilitarian tradition does not provide any ground for determining how this number is to be arrived at. Perhaps a general ground should not be sought. Perhaps we should sum over all those questions that arise within the lifetime of a particular parliament, for example. There are great difficulties here, and I cannot see that in principle the decision can be made, other than on a philosophically arbitrary basis. But this criticism does not lie just against this 'fairer' form of democracy. I wish to stress that it appears equally arbitrary how we delineate that population of which a majority is to be taken. No standard for this is provided within the liberal utilitarian tradition, and the problems which arise as a result are particularly obvious here in Northern Ireland.

Finally, Miss Anscombe remarked on a possible danger: it will, she says, be possible for the tyrant to damage the interests of any group while truthfully claiming democratic support for his measures.

It is true that, if a group is in a minority on every issue, then it will be damaged on every issue if the decision is in accordance with the wishes of the majority. This is a well-known feature of democracy:

majorities winning implies minorities losing. This matter does not arise peculiarly from Miss Anscombe's argument.

Miss Anscombe seems to be claiming more than this trivial fact. It appears to worry her that damage to interests may be occurring haphazardly now for all we know. But this is to do no more than re-state her description of her table: the majority may vote in the minority on the majority of the questions.

I believe that Miss Anscombe is making something like the following claim: that her table shows that a tyrant could attack some sectional interest which, in some hidden way, is shared by the majority of the people. But this is not possible. For a sectional interest is an association of desires, and only incidentally of people; and we have seen that where questions are answered by majority decision the majority of desires will always be met. But, although I may not have exhausted all the claims that Miss Anscombe may be making in her expression of alleged danger, a quite general argument is available to refute this danger. The outcome of a series of questions is determined by the actual desires people have and their distribution within a population, and, on the present assumptions, all that an aspiring tyrant can do is note how the dice of desires fall.

The Queen's University of Belfast   © J. L. GORMAN 1978

RUDINOW AND SIKORA ON ART-CRITICAL CONCEPTS

By Henning Jensen

IN a provocative article 'Are There Art-Critical Concepts?'[1] Joel Rudinow and Richard I. Sikora argue against what they refer to as the non-conceptual or, for short, NC thesis, the thesis which denies that art-critical terms correspond to art-critical concepts and which, they maintain, comes to this: 'In art criticism one does not, strictly speaking, describe the work of art; those of its features which are the particular concern of the critic (let us say its aesthetic features) are not shareable' (p. 196). Their arguments against the NC thesis rest on their use of the distinction between determinate and determinable features. Specifically, they maintain that the point of art criticism is to draw attention to a work's non-shareable determinate features by describing it in terms of its shareable sub-determinable features. And by insisting that in art criticism the work is, strictly speaking, described, they claim to have indicated a way of avoiding the NC thesis. I shall argue, to the contrary, that the distinction between determinate and determinable features provides no support whatsoever for their case against the NC thesis and, furthermore, that, far from avoiding it, they have defended a position which is not only consistent with this thesis, but nearly identical with it.

A brief summary will serve to outline the position defended by Rudinow and Sikora and show how it relates to the NC thesis. Rudinow and Sikora grant that there are peculiarities about art criticism which render the NC thesis attractive, peculiarities such as the use of figurative language and the dependence on first-hand familiarity with the work being criticized. However, they find this thesis counter-intuitive in its claim that art-critical terms change meaning from employment to employment and in the ensuing puzzle as to how, in the absence of fixed meanings, one could talk of meaning and communication in art criticism. To avoid these counter-intuitive characteristics of the NC thesis Rudinow and Sikora argue that there are sub-determinable art-critical concepts which do, strictly speaking, describe the work of art. To account for the characteristics which render it attractive, they suggest that the point of such descriptions is to draw attention to the non-shareable determinate aesthetic features of the work of art.

I now want to examine the arguments which Rudinow and Sikora present against the NC thesis. Since they maintain that the distinction between determinates and determinables provides a basis for the "finessing" of the NC thesis, it might be best to begin by examining this distinction. But before turning to their treatment of this distinction, I want to make some preliminary observations of my own. Current use of

the terms 'determinate' and 'determinable' may be traced to W. E. Johnson who said: 'I propose to call such terms as colour and shape *determinables* in relation to such terms as red and circular which will be called determinates . . .'[1] The distinction may be used both to classify terms like 'red' and 'scarlet' and the properties and qualities named by them. However, it may be remarked that if we are classifying terms, no term such as 'scarlet' applies to an absolutely specific colour. Johnson suggests that this practical impossibility in the use of terms does not prevent us from postulating the absolutely determinate characters of things. The distinction between determinates and determinables, therefore, may be seen to concern the relations between the more or less specific. But it is of considerable logical interest to note that unlike the relation of genus to species in which species must be marked off by differentiating properties, the determinate is marked off within the determinable, as Searle so happily phrases it, 'without outside help'.[2] It is also of logical interest to note how the distinction between determinates and determinables bears on the distinction between the general and the particular. Now it so happens that the term 'particular' is notoriously ambiguous. It may mean 'specific' as opposed to 'general' or it may be used in the sense in which 'particular' is contrasted with 'universal' and in which to be a particular is to be an instance. It is of great importance to my later arguments against Rudinow and Sikora to insist that the distinction between determinates and determinables concerns relations between the particular and the general where the term 'particular' is used in the former of the above senses, but must not be confused with relations between the particular and the universal where the term 'particular' is used in the latter of the above senses.[3] Thus the particular, again in the latter of the above senses, must not be confused with the determinate, even the *absolutely* determinate. Whereas to be a particular in this latter sense is to be an instance and to be incapable of being shared, the notion, for example, of an absolutely specific or determinate shade of red is the notion of a property which is abstract and might be shared by other occurrences of this colour.

Turning now to Rudinow and Sikora, we find them drawing the following distinction:

> A determinate feature is such that two or more objects which have it must be in that respect qualitatively indistinguishable. A (sub-)determinable feature is such that two or more objects which have it need not be in that respect qualitatively indistinguishable (p. 198).

They now proceed to point out that in talking about an object's

---

[1] W. E. Johnson, *Logic* (Cambridge: Cambridge University Press, 1921) Part I, p. 174.
[2] S. Körner and John R. Searle, 'On Determinables and Resemblance: Symposium', *The Aristotelian Society*, Supp. Vol. XXXIII, 1959, p. 143.
[3] Cf. L. S. Stebbing, *A Modern Introduction to Logic* (6th ed.; New York: The Humanities Press, 1948), p. 444 and R. M. Hare, *Essays on the Moral Concepts* (London: Macmillan, 1972), p. 27.

determinable colours one is describing it in terms of shareable features. However, they suggest that perhaps a work of art's determinate aesthetic features are not shareable. But I should argue that the claim that the work of art's determinate aesthetic features are not shareable is self-contradictory, not only in the light of my earlier remarks, but even according to the position outlined thus far by Rudinow and Sikora. In my earlier remarks I established that both determinate and determinable features are shareable, involving abstraction and the use of concepts. And in the above quotation Rudinow and Sikora include the words 'a determinate feature is such that *two or more objects which have it . . .*' (italics added). In this passage, therefore, they establish that to be a determinate feature is to be shareable, thereby contradicting their later suggestion that determinate aesthetic features are not shareable. One may suspect that they are led to embrace this incoherent position by succumbing to the all too familiar error, described above, of mistaking determinateness for particularity where the latter concerns the sense in which to be a particular is to be an instance. In short, their use of the distinction between determinates and determinables leads them to make the incoherent claim that there are determinate aesthetic features which are non-shareable and, in general, provides no support whatsoever for their case against the NC thesis.

Rudinow and Sikora conclude their article with suggestions concerning a position which, they claim, will avoid the NC thesis while doing justice to those features of art criticism which render it attractive. Briefly stated, their position is this: (1) the point of art criticism is to draw attention to the determinate non-shareable features of works of art; (2) attention is drawn to the latter features in a round-about way by describing the work in terms of its shareable sub-determinable features. I now want to compare this position with that of defenders of the NC thesis in order to discover whether Rudinow and Sikora really avoid this thesis and offer a defensible alternative to it. While making this comparison I propose to leave aside considerations concerning the use which they make of the distinction between determinates and determinables and which I criticized above. To begin with, it is most important to emphasize that in allowing for the presence of non-shareable aesthetic features and in maintaining that the point of art criticism is to draw attention to these features and not to the similarities between works of art, the position defended by Rudinow and Sikora is indistinguishable from that of defenders of the NC thesis such as Hampshire, Strawson, and Isenberg. Hampshire, for example, states that '[the critic's] purpose is to lead people . . . to look here, at precisely this unique object; not to see the object as one of a kind, but to see it as individual and unrepeatable'.[1]

---

[1] Stuart Hampshire, 'Logic and Appreciation', in William Elton, ed., *Aesthetics and Language* (Oxford, Basil Blackwell: 1954, p. 165).

Strawson, expanding points made by Hampshire, recommends the con-
trasting of what he calls 'aesthetically relevant *features* of works of art'
with what he calls 'shareable properties of works of art'.[1] And I shall
assume that Rudinow and Sikora have established that, as regards the
point of art criticism, Isenberg also subscribes to the NC thesis.

Since Rudinow and Sikora are clearly in agreement with the NC
thesis as regards the ultimate point of art criticism, their claim to have
avoided it must therefore involve some disagreement with this thesis
with respect to the means through which the point of art criticism is
attained. And the source of this disagreement is not hard to find. Rudi-
now and Sikora maintain that the point of art criticism is attained by
means of descriptions of the work in terms of its shareable features.
Hence, their claim to have avoided the NC thesis must follow from their
taking this thesis to mean that description plays no role whatsoever in
art criticism, that descriptions are involved neither in the point of art
criticism nor in the means by which this point is attained.

But are Rudinow and Sikora correct in supposing that the NC thesis
must mean that description plays no role whatsoever in art criticism? In
addressing this question, I now want to show that the position taken by
Rudinow and Sikora with respect to art-critical concepts and the role of
description in art criticism is, contrary to their contentions, entirely
consistent with the NC thesis. To begin with, it should be insisted that
the term 'NC thesis' is dangerously misleading if it is taken to mean
that concepts have no place whatsoever in art criticism. The latter view
is utterly false. How, it might well be asked, could one possibly indulge
in art-critical talk without using concepts at all? It is worth noting that
Isenberg, identified by Rudinow and Sikora as defending the NC thesis,
devotes a good deal of attention to the role of concepts in art criticism.
Far from downgrading the role of concepts, he takes great pains to dis-
sociate his position from that of such writers as Bergson and Croce
whom he reports as saying that 'it is impossible by the use of concepts to
"grasp the essence" of the artistic fact'.[2] His own view may be sum-
marized by attending to what he says about the purpose of criticism and
how this purpose is served by concepts. Of the purpose of criticism he
says: 'it is a function of criticism to bring about communication at the
level of the senses, that is, to induce sameness of vision, of experienced
content' (ibid., p. 163). And concerning concepts and criticism he re-
marks on 'the actual and very large influence of concepts on the process
of perception . . .' (ibid., p. 167). It seems evident that while defending
the NC thesis, Isenberg may nevertheless agree with Rudinow and
Sikora in claiming that concepts are used in art criticism.

[1] P. F. Strawson, 'Aesthetic Appraisal and Works of Art', The Oxford Review, 1965,
p. 13.
[2] *Aesthetics and the Theory of Criticism: Selected Writings of Arnold Isenberg* (Chicago:
University of Chicago Press, 1973), p. 165.

The only issue remaining, therefore, would appear to involve the contention by Rudinow and Sikora that although the point of art criticism is to draw attention to the non-shareable features of the work of art, this point is nevertheless attained through what are, strictly speaking, descriptions. But now a close examination of Isenberg's position indicates that he would not disagree. Consider, for example, the following typical sentences from Isenberg: '. . . the critic's *meaning* is "filled in", "rounded out", or "completed" by the act of perception, which is performed not to judge the truth of his description but in a certain sense to *understand* it'; 'Every descriptive statement affects our perception of— and our feeling for—the work as a whole'. From these sentences and from what we found him saying about concepts, it would seem that he might well maintain, like Rudinow and Sikora, that, strictly speaking, concepts and descriptions are employed in art criticism. At the same time, of course, he will maintain that, as regards the point of criticism, 'the *meaning* of a word like "assonance"—the quality which it leads our perception to discriminate in one poem or another— is in critical usage never twice the same' (ibid., p. 165). I conclude, therefore, that there is no important difference between Isenberg's position and that of Rudinow and Sikora.

How did Rudinow and Sikora come to suppose otherwise? Let me try to identify the fatal flaws in their arguments. The most fundamental error to be found in their position may be explained as follows. Their claim to have avoided the NC thesis rests on the contention that they have maintained what the NC thesis denies, namely, that in art criticism the work is, strictly speaking, described. But upon careful examination it becomes clear that in the course of their article references to the strict use of describing occur in two contexts. Thus when they say 'the NC thesis comes to this: In art criticism one does not, strictly speaking, describe the work of art' (p. 196), the issue surely concerns the point of criticism or its defining characteristics. Whereas when they defend their own view that in art criticism the work is, strictly speaking, described, the issue concerns the use of description as a round-about way of attaining the point of art criticism. Therefore, the claims of Rudinow and Sikora concerning the latter use of description do not contradict the claim of the NC thesis about the very point of criticism. In addition, as my quotations from Isenberg have shown, they are quite in error in ascribing to defenders of the NC thesis the view that concepts and descriptions can play no role whatsoever in art criticism.

In summary, I have argued, first, that, contrary to Rudinow and Sikora, the distinction between determinate and determinable features provides no way of finessing the NC thesis. Second, I have tried to show that, far from avoiding the NC thesis, they have defended a view which is not only consistent with this thesis, but nearly identical with it.

Finally, I should want to insist that my discussion leaves as an open question the truth or falsity of the NC thesis itself or, as I should prefer to say, of those views which stress the non-conceptual character of art criticism.

University of Arizona   © HENNING JENSEN 1978

ASSERTIONS: A REPLY TO COHEN

By D. H. M. BROOKS

A MODEL of a bivalent language L can be seen abstractly as an ordered triple, $<\{0, 1\}, \{1\}, F>$ where F is a valuation function mapping sentences of the language onto the values 0 and 1, and $\{1\}$ is the set of designated values. A sentence s of L is generally said to be true iff $F(s) = 1$. In a bivalent language, such as L, F can be defined by recursively specifying for each sentence s the conditions on which it is assigned the value 1, and saying that if these conditions are not fulfilled then s is assigned the value 0. To specify in detail the conditions on which s is assigned the value 0 would be wasted effort. (With a trivalent language, however, to specify only the conditions on which a sentence was assigned the value 1 would be inadequate, for in the case where these conditions were not fulfilled it would be left undecided whether a sentence had the value 0 or 2). It is only for a bivalent language, then, that the valuation function in its model can be defined by picking out one of the possible values and specifying for each of the sentences of the language the conditions on which it has that value, noting that if these conditions are not fulfilled the sentence has the other value. When such a definition is given, the value picked out (designated) is generally identified with truth and the other value with falsity. F can, however, be defined just as adequately by picking out the undesignated value 0, and recursively specifying for each s of L the conditions on which s has the value 0 and stating that s has the value 1 otherwise. We might call this a falsity definition for L and the other specification of F a truth definition for L.

In ANALYSIS 36.3, pp. 113–17, I described a situation which could be characterized thus; two languages *Mend* and *Ver* are spoken on an island and are such that the truth definition for *Ver* is the same as the falsity definition for *Mend* and vice versa. More formally the interpretation $<\{0, 1\}, F>$, where F is defined by recursively specifying for each sentence *both* the conditions on which it has the value 1 *and* the conditions

on which it has the value o, serves equally well for both languages. The languages differ only in that speakers of the one attempt to utter only sentences with value o, while speakers of the other attempt to utter only sentences with value 1. I suggested that such an interpretation provided one account of two languages in terms of the sense and reference of their sentences.

As Michael Cohen points out in ANALYSIS 37.1, pp. 44–5, this suggestion is mistaken. While our characterization of F is the same for both languages and might contain sentences such as these:

(i) $F(\varphi a) = 1$ iff a is round,

(ii) $F(\varphi a) = 0$ iff a is not round,

since speakers of *Ver* attempt to utter only sentences assigned the value 1 and speakers of *Mend* attempt to utter only sentences assigned the value o, in *Ver* 'φ' means *round* and in *Mend* 'φ' means *not round*. Obviously 'φ' does not have the same sense in both languages and the account I have outlined cannot be said to characterize the sense of expressions in a language. (This assumes that the hypothesis that speakers of *Mend*, say, are attempting to utter false sentences is ruled out. Cohen considers this hypothesis incoherent. I am not sure, though, that it might not be the correct way to characterize the first generation of a perverse religious sect who deify falsity, and accompany each utterance of a false sentence with the thought 'and I dedicate this falsehood to you, O falsity'. Consistency in their worship, however, will prevent them from passing down the full flavour of their belief to their children since the beliefs that the rest of society is perverse etc. will change from leaps of faith to commonplaces. This is perhaps the sole instance of long acceptance of a religious dogma being evidence for its truth, in that it, in fact, makes the dogma true.)

It could however be argued that a model of a language of the sort I have outlined does provide an account of sense in terms of truth conditions. What after all in a bivalent logic do we mean by designated and undesignated values if not truth and falsity? Do we not here call truth and falsity designated and undesignated values only because two valued logic is the limiting case of a many valued logic? Where else can we speak of truth if not here? The model, itself, provides the best definition of truth one can get. In defining the function F we define truth by specifying its extension in L. Every instance of the schema '*s*' *is true iff s* is derivable from our model. We have in Tarski's sense defined truth.

This argument fails, because, just as the model is inadequate as an account of sense, it is inadequate as an account of truth. Tarski's definition only becomes a definition of truth, when supplemented with the information that users of the language for which truth is defined attempt

to utter only true sentences. So too, if sense is to be given in terms of truth conditions, and if my schema for such an account is correct, then supplementing the model with the information that speakers of the language attempt to utter only sentences with designated value converts it, for indicative sentences, into an adequate account of sense in terms of truth conditions.

This fundamental point will affect any attempt to explain meaning in terms of truth. Whatever form such an account takes there will be an isomorphic account differing from that account only in that it is framed using a purely semantic notion such as that of having a designated value, which, while it is a large part of the concept of truth, is not the concept of truth. Labelling the designated and undesignated values of a formal semantics 'true' and 'false' only obscures this point. Insofar as it seems to make the account adequate it does so by inducing us to read into the metalanguage a non-formal empirical fact about the way the object language is used. This reading in makes Tarski's theory of truth seem more plausible than it in fact is. (These different ways of labelling the designated and undesignated values account for the different ways in which Dummett approaches his analogy between chess and language, which Cohen picks out in ANALYSIS 36.1, pp. 2–4.)

An account of a language in terms of the sense and reference of its expressions is adequate. No purely formal account of a language can be adequate unless supplemented by the information that speakers of the language attempt to utter only sentences belonging to a particular class. This gives us an account of the sense of indicative sentences in terms of truth conditions. Supplementing the formal account with, for example, a specification of the language's grammatical device for transforming indicative sentences into imperative sentences and the information that, roughly, utterers of imperative sentences intend that their hearers should bring it about that the corresponding indicative sentence has a designated value (falls into the class of sentences speakers of the language attempt to utter), gives us an adequate account of the sense of imperative sentences. If, as Cohen says, it is a 'rather trivial fact that one (can) not be said to understand a command unless one also understand(s) the (corresponding) assertion' (ANALYSIS 37.1, p. 45) then the fact that one can understand either and hence communicate at all would seem to be equally trivial.

University of Cape Town © D. H. M. BROOKS 1978

CACODAEMONY AND DEVILISH ISOMORPHISM

By JOHN KING-FARLOW

IN "Cacodaemony" (ANALYSIS 37.2) Steven Cahn claims (A) that little or no discussion has been devoted to 'the problem of good', the problem of deciding whether 'a world containing goodness could have been created by an omnipotent, omniscient, omnimalevolent being' (p. 69). He adds (B) that 'all the arguments and counterarguments [applied to the traditional problem of evil] are applicable *mutatis mutandis*' to his problem of good (p. 72); (C) that John Hick's famous theodicy can be paralleled by this equally strong cacodaemony (p. 72); (D) that theist and demonist assertions must therefore be equally *improbable* (p. 73); (E) that, unless 'demonists or theists can produce other evidence in favour of their views', then 'the reasonable conclusion is that neither the Demon nor God exists' (p. 73).

A few comments and reminders seem to be in order. (A¹) This sort of *argument from good* was used quite recently by Edward Madden and Peter Hare in their book *Evil and the Concept of God* (Springfield, Illinois, 1968): the work was compiled from earlier articles on theodicy in several journals, including one on this problem of good in *Philosophy and Phenomenological Research* 1967. In the book we read, for example: 'the problems of evil and good are completely isomorphic . . . if one is insoluble so is the other.' (p. 34). In the book (pp. 3–4) and in *Philosophy and Phenomenological Research* 1969 they adopt a conclusion like Cahn's on the matter of probability.

(B¹) Today the friends and the foes of theodicy sometimes seem to forget that in the relevant tradition the primary goal of answering arguments from evil is much more like the purpose of a logician trying to establish the independence and consistency of various propositions, much less like the purpose of a scientist proving a hypothesis' truth or establishing a high probability for it. Thus 'consistency proofs' for cacodaemonism, Manicheanism, henotheism, polytheism and even some godless beliefs in justice through transmigration would not impugn the consistency of monotheism any more than they would that of Euclid. Rather, they might show how the possible addition of certain ideas outside the Judaeo-Christian tradition might clarify the coherence of monotheism.

(C¹) From the standpoint of actual people engaged in practical reasoning about *commitment* to a religious or ideological standpoint, claims about the equal consistency and equal probability of theist and demonist groups of possible beliefs are likely to turn out to be ludicrously irrelevant. Most individuals who are seriously considering such a new commitment—in countries where ANALYSIS is ever read—are concerned with sets of what

William James called *live hypotheses*. These have come to form *momentous* or *forced options* for those people in their historical context. It is sets like {*theism, agnosticism, naturalism*}, {*traditionalism, nihilism*}, {*Catholicism, Maoism*}, {*liberalism, 'apolitical' anarchism, Neo-Marxism*} which many highly literate, informed, reflective, and careful people really contemplate after a great deal of living in the world. Not even the recent popularity of films and books like *The Exorcist* encourages them very often to consider such Hothouse Cacodaemonizing.

(D¹) Cahn, like Madden and Hare, would have us believe that their devilish isomorphisms demonstrate this: all reasonable people ought to take demonism as seriously as theism, hence they should take both very lightly. This move might be more plausible IF profound historical surveys of apparently 'theist' and 'diabolical' *experiences* could be made, and IF all reasonable people could thence know that the surveys produced another devilishly tied contest. Of course, no such thing is available. But, anyway, it is a strange demand on human rationality to expect each reasoner to shuck off his or her own historical experience and feel equally interested in any new, logically possible Ism. To make such a demand is to beg the question that such reasoners are not entitled to regard the very fact that they are each born into a peculiar place in history as an important indication of what they each ought to be considering seriously as religious or ideological options.

(E¹) 'The reasonable conclusion is that neither the Demon nor God exists' (Cahn, p. 73). In other words, (i) such an isomorphism gravely affects the *probability* of the existence of either; (ii) therefore, the reasonable person should reject both. As for (i), this involves a dubious generalization from an alleged draw between one theodicy and one cacodaemony to faith in the perfect isomorphism of all possible pairs of variations on those themes. To be greatly impressed by (ii), one would have to think that no progress had been made on inductive logic and practical reasoning since Pascal's first birthday. In the making of rational decisions in the face of radical uncertainty, the *probability* of a proposition may be quite low, so long as it is positive, and yet not lose its interest for the reasonable chooser. For some sufficiently high and ethically appropriate *utility* of accepting the proposition and being right, together with a severe disutility of rejecting it and being wrong, may well make it the *reasonable conclusion* for a wise person that he or she should be systematically committed to beliefs and deeds which flow from such a proposition. The cerebral purity of philosophers who eschew thoughts of appropriate utilities in favour of probability alone, when discussing religious and other deeply practical issues, is the purity of a spotlessly gleaming and perfect mirage. The rational man who searches for living water in the desert will prefer an imperfect, dusty and otherwise untidy oasis. A risky wager to *Maximize Expected Utilities* of ethically promising

kinds (in the face of radical uncertainty) can at least be part of a realistic strategy. Even if Cahn were right about the relevant probabilities, his conclusion would be a bizarre place for putting an end to concerned reasoning about a religious enigma.[1]

University of Alberta   © JOHN KING-FARLOW 1978

[1] In ANALYSIS 30.4 (pp. 140–4) I discussed the one-sidedness of Madden and Hare on matters now raised by Cahn; some simple implications of elementary Decision Theory for practical reasoning were spelled out. In *Faith and the Life of Reason* (Reidel: Dordrecht, 1972) William Christensen and I developed these implications so as to relate them to Pascal and James. We also replied there to Lawrence Resnick's ingenious attack on that 1970 paper, in his 'Evidence, Utility and God', ANALYSIS 31.2 (pp. 87–90).

THE DISTRIBUTION GAME

By HILLEL STEINER

THREE persons materialize from out of nowhere on an otherwise depopulated Earth. Before them stands a vast department store. They are addressed by a booming voice which emanates from every-where. The voice announces:

> Henceforth, enforced conduct shall be governed by, and only by, one principle: *You own what you earn.* To *own* something is to be en-titled to the noninterference of others with any use to which you choose to put that thing. To *earn* something is either (i) to be the first possessor of that thing, or (ii) to be someone to whom the owner of that thing has chosen to transfer the ownership of it, or (iii) to be the owner of all the things which, when synthesized, produced that thing. Neither first possession nor transference nor production count as earning when performance of these activities is secured by interference with others or with what they own.

The voice continues:

> Any object, the acquisition of which is attempted by means other than earning, will immediately be rendered physically unusable for its would-be acquirer alone.

The voice concludes:

> You see before you a vast department store. In it are housed the title deeds for each and every object in the world. To be the first to hold the title deed to an object is to be its first possessor. Upon completion of the sentence I am now uttering, the doors of the store will unlock and you may enter the store and take possession of these deeds.
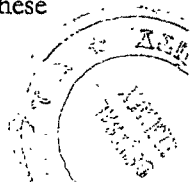
The voice stops speaking, the doors unlock, the three persons enter and, between them, establish their ownership of each and every object in the world.

The next day, three more persons materialize from out of nowhere on an Earth now populated by the three persons who preceded them. They are addressed by the voice which again announces the principle stated above, explicates it, and informs them of the manner in which it is enforced. The voice does not, however, add the concluding remarks which it had addressed to the new arrivals' predecessors. It does not do so because, although they too are standing before the same department store, both it and the ground they are standing on—as well as everything else in the world—are now owned by one or another of the original three. And indeed, even before the voice has finished speaking, the new arrivals experience rapidly mounting difficulty in standing on that ground and in breathing the surrounding air. However, just as they are about to succumb to the fatal effects of these natural enforcement agencies, they see a person (the owner of the land and air-space they are occupying) approaching them with a pen in one hand and a labour contract in the other.

Is the ensuing social order one which conforms to the principle 'You own what you earn'?

University of Manchester   © HILLEL STEINER 1978

UNFAIR TO GROUPS: A REPLY TO KLEINBERG

By PAUL WOODRUFF

STANLEY S. KLEINBERG argues against my suggestion that discrimination, *when wrong*, is unfair to groups (ANALYSIS 37.1, pp. 46–48). His objections stem from a certain looseness of expression in my original statement (ANALYSIS 36.3, pp. 158–60), which I now propose to correct.

Kleinberg uses 'discrimination' to apply only to practices that are wrong; I use it of certain practices whether or not they are wrong. I do so in order not to beg the question whether compensatory discrimination is in principle wrong. So that we may state the issue between us, let us agree that discrimination in hiring is denying jobs to applicants because they belong to certain groups, when membership in those groups is irrelevant to qualifications for the jobs. That is not adequate for identifying cases of discrimination ('irrelevant', 'qualification', and the sense of 'because' need to be explicated); but it allows me to state the differences between Kleinberg and me. (1) Kleinberg holds, as I do not, that discrimination in hiring violates an obligation to be fair. (2) I hold, as

Kleinberg does not, that discrimination in hiring is unfair only when it is unfair to a group.

Kleinberg appears to rely on current usage. Certainly in cases of hiring (unlike wine-tasting) we allege discrimination only as a complaint. But that will not settle the issue of whether the practice I described violates an obligation to be fair. We need to know when such a complaint is justified, and on what grounds. I argued from an example that discrimination in hiring is not always unfair. I went on to suggest that when it is unfair, it is so because it is part of a pattern that unfairly reduces the respect in which a group is generally held. That cannot be a complete account of the matter. I did not say what it is for a group to have or to lose respect, or how much respect it is fair for a group to have; nor did I specify the principle of fairness involved. To complete the account would require a substantial essay, not suitable for this format. Let us limit our discussion to cases involving racial groups. For them the principle of fairness of respect is simply equality; such groups, I propose, should receive equal respect, and be treated by society in whatever way is required to bring them equal respect. I said that when a group is made to receive less than its fair share of respect it is *insulted*. When a group is insulted in this technical sense, all its members lose respect unfairly just in so far as they are thought of as members of the group. When an individual is denied a job because of his membership in a group, whether or not he is personally insulted depends on contingencies irrelevant to this discussion. For discrimination is essentially *im*-personal; what the discriminator has in mind is groups, not persons. Whether or not a *group* is insulted by an act of discrimination depends on two things. First, the act cannot be private and isolated; it must belong to a pattern if it is to affect the respect in which the whole group is generally held. Second, the effect on the group must be unfair; the discrimination must contribute either to reducing the group's respect below a fair level or, if that is already down, to maintaining it there. It follows that discriminating against a group with more than its fair share of respect is not insulting in the technical sense to the group. If its members suffer hurt feelings, that is another matter.

The fact that certain groups have been insulted by discrimination in hiring is beyond dispute. But compensatory discrimination is often supposed to be *prima facie* unfair because it is discrimination. If I am right, that objection fails. If discrimination is wrong only when it insults groups, it is not wrong in every case. Kleinberg thinks I am not entitled to that conclusion. 'In order to avoid self-contradiction [Woodruff] must be construed as holding that particular types of behaviour can be saved from being insulting by virtue of not being part of a pattern' (p. 46); but, Kleinberg holds, an insult is an insult whether or not it is part of a pattern. This objection plays on the incompleteness of 'insulting'

(insulting to a group or a person?) and does not tell against my claim that an isolated act of discrimination does not insult a group. Insults to groups can be accomplished by private acts only in patterns. Kleinberg's objection on this score is at best a *petitio*, resting on the concealed assumption that whatever is unfair about discrimination is unfair directly to the persons concerned, and thereby indirectly to their group. But that is the point at issue. I deny that discrimination in hiring is unfair directly to the individuals involved.

Kleinberg raises two questions about compensatory discrimination (p. 47). (1) Does it provide a benefit that is distributable to members of the group? (2) Can it compensate for discrimination or (as Kleinberg asserts) at most end a pattern of discrimination? On the first point I think the question empirical whether compensatory discrimination can raise the level of respect in which a group is generally held. If it can, its benefits are distributable; for belonging to a properly respected group is better than belonging to an insulted one. On the second point the issue is partly empirical as well. But Kleinberg's assertion raises the question of how 'compensatory' in 'compensatory discrimination' is to be understood. There are two ways of compensating for a practice that leads to inequality. The first compensates for existing prejudice by introducing a practice that exerts, so to speak, a compensating pressure, so that equilibrium results. The second requires those responsible for the bad practice in the past to make restitution to the persons they have injured. Discriminating in favour of an insulted group might well accomplish the first kind of compensation. But it could not achieve restitution. Restitution is required of the guilty, but discrimination works against entire groups. That is why compensatory discrimination carries an air of injustice; it appears to exact payment from the wrong people. But if compensatory discrimination is confined to the goal of equilibrium, it is not unjust in that way. If some members of a previously favoured group encounter discrimination and are unemployed, that is hard for them. But unemployment is an evil quite apart from discrimination, and I do not see how it could be fairly distributed in a society that values jobs. So if these new unemployed are entitled to complain of unfair treatment, they should complain of unemployment, not discrimination. It is one of the consequences of past discrimination (and the sort of thing compensatory discrimination seeks to correct) that complaints from a highly respected group are more likely to be heeded.

One small point. Kleinberg misquotes me on p. 47. I did not say that discrimination was 'according differential treatment on morally irrelevant grounds'. In fact, I argued against the relevance of moral relevance in this context (p. 159, para. 2).

The University of Texas at Austin   © PAUL WOODRUFF 1978

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Christopher Kirwan, Exeter College, Oxford OX1 3DP. Contributors are asked to note the following requirements.

Articles should normally not exceed 3,000 words in length. Occasionally, however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope and international reply coupons of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

ANALYSIS "PROBLEM" NO. 17

THE seventeenth problem is set by Professor G. E. Hughes of the
Victoria University of Wellington, and is as follows:

> Can I ever, by my subsequent actions, bring it about that something
> I did on a previous occasion was done from a certain motive rather than
> from some other one?

The word limit is 600 words. Entries should reach the Editor of
ANALYSIS by 31 August 1978; they should not be sent to Professor
Hughes. Entries will not be acknowledged or returned unless accom-
panied by stamps or international postage coupons. Contributors may
submit entries under their own names or a pseudonym. Contributors
must be under the age of thirty, or undergraduates or graduate students.

A report with any winning entries will be published in volume 39
of ANALYSIS. The ANALYSIS Committee has voted a sum of £40 which
will be awarded as a prize if Professor Hughes finds a sufficiently
deserving contribution.

The report on Problem No. 16 will appear in the next issue.

THE POWER OF RUSSELL'S CRITICISM OF FREGE: 'ON DENOTING' pp. 48-50

By SIMON BLACKBURN *and* ALAN CODE

1. SETTING the Scene.

In 1905 Russell first published, in 'On Denoting', the theory of
descriptions. He was conscious at the time that the theory made a pro-
found break from its predecessors, the theories of Meinong, of Frege,
and of his own work of two years earlier. Thus he included in his
article a discussion of the weaknesses of those theories. His readers have
had no difficulty in understanding just what his criticisms of the Mein-
ongian theory are—indeed for a very long time it was generally conceded
that he had successfully undermined it. His criticisms of Frege, and of
his own earlier theory which he regarded as similar, were not blessed
with such a fortunate reception. To the best of our belief only one
author, A. J. Ayer, apart from Russell himself, has ever acknowledged
in print that Russell had an argument which seriously threatened Frege's
view. It is usually held that either he had no argument at all against any
previous theory, because he was hopelessly muddled, or that at best he

had succeeded in undoing his previous view, expounded in *The Principles of Mathematics* [10], but totally failed to engage Frege. Dummett thinks that Russell does, in a confused fashion, manage to make a minor point which Frege can easily evade. These assessments are quite wrong. We do not say, on this occasion, that Russell had a decisive, or overwhelming, objection to Frege. But we do say that he presented a serious argument and even "considerations capable of determining the intellect" against Frege's theory. To prove this we offer a rational reconstruction of the train of thought found in pp. 48–50 of 'On Denoting', as numbered in the collection *Logic and Knowledge*, edited by Robert Marsh[9]. For convenience we shall letter the relevant eight paragraphs by the italic capitals (*A*) to (*H*), and refer to them in this way. (*A*) begins at the bottom of p. 48 (The relation of the meaning to . . .) and (*H*) at the bottom of p. 50 (That the meaning . . .).

The issues are bedevilled by problems of terminology. We shall call both proper names and definite descriptions 'denoting phrases'. These are the things which are in dispute. We shall talk of them having a 'sense' and a 'reference'. The reference of the word 'Aristotle' is Aristotle, and the word refers to Aristotle. The sense of the word is supposed to be a third entity. It is not Aristotle, nor is it the word 'Aristotle', but instead is some entity associated with the word 'wherein the mode of presentation is contained' [6] p. 57. Part of Russell's point will be the difficulty of referring to such a thing, so we musťbe pardoned for saying no more about it. We shall call the view that denoting phrases have usually both sense and reference the three-entity view, for obvious reasons. We also need a way of describing the relationships that hold between these three things. We shall say that a word 'refers to' its reference and 'expresses' its sense. If a sense relates to a thing in such a way that a word which expresses it refers to that thing, or in such a way that a word which were to express it would refer to that thing, we shall say that the sense 'determines' the thing. Of course further theses about all these things would be needed to describe fully anyone's theory, but for the present we merely introduce the terminology.

2. Previous Interpretations.

As we stand nearly alone in our insistence that Russell was at least in the right ball-park, we would like to say a few words about the more usual views.

Church [4] p. 302, and Butler [2] pp. 361–363 hold that Russell's attempt to refute Frege is vitiated by his (characteristic) failure to observe the use/mention distinction. On this view, once some consistent use is made of perspicuous notational devices distinguishing talk of expressions from talk of senses of expressions, and both from talk of references of expressions, then the alleged argument vanishes. Now Cassin [2]

p. 269, has shown that there is no way of interpreting Russell if Church and Butler are right. And on the face of it the view is implausible. While it is undeniable that Russell is quite careless in his notation, he was very sensitive to the underlying point, which, indeed, functions centrally in his own argument. Paragraph (D) is mainly devoted to explaining it. Since Russell gave up his own earlier position because of the present argument, Church and Butler put him in the following ridiculous position. There is an argument which purports to refute Russell's old view, and which emphasizes and explains the use/mention distinction. But the argument disappears once this distinction is made. Nevertheless Russell feels forced to abandon his old view. Clearly we should not accept this exegesis unless absolutely no other is forthcoming: it offends against the cardinal principle of Russellian exposition, to wit: don't make him out a complete fool if you can help it.

More common is the view that Russell did have some sort of argument going, only it bears on his earlier position, not on Frege, who only gets dragged in because of Russell's notorious inability to read properly. This is the view expressed in Geach [7]. As he there puts it: "readers of 'On Denoting' will find it best simply to ignore his use of Frege's name". To approach this we must have some idea of how the two theories differ—Geach does not tell us which feature of Russell's earlier theory renders it, but not Frege, a possible target for "On Denoting".

There are two important differences between the theory of *The Principles of Mathematics* (which we refer to as *PoM*) and Frege. The first, however, turns out to be basically terminological, and the second cannot possibly matter to Russell's argument. The first is that for Russell it is concepts which denote, so that it is the denoting concept meant by a definite description which denotes the thing referred to, or denotation. On Frege's theory it is the definite description itself—the actual expression—which refers to the reference. Thus we cannot have all of: sense=meaning; reference=denotation; referring=denoting; for this does not map Russell's theory onto Frege's. In Frege one kind of thing (an expression) refers, and in Russell a different kind of thing (a concept) denotes. But this is trivial by itself, since, if there is a relation which obtains between the meaning of the description and the denotation, we can define in terms of it another relation holding between the description itself (the words) and the denotation of its associated meaning. And likewise if there is a relation which obtains between a word, or group of words, and its reference, we may define in terms of that relation another one holding between the sense of the expression and that reference. Using our triangle of terms we can introduce 're-ferring' once we have 'expressing' and 'determining', or we can introduce 'determining' once we have 'expressing' and 'referring'. The former order of definition, Russell's, seems preferable, since in Russell and Frege

a word has the reference it does as a consequence of having a certain sense, and of this sense determining a certain thing. This point is important to Russell later, as we shall see in section 3, but for the present it is clear that the difference of terminology marks no respect in which his own theory is less acceptable than Frege's.

The second difference is that for Frege the sense/reference doctrine applies to every singular term, whereas at the time of *PoM* Russell applied his analysis (into denoting concept/object indicated) only to definite descriptions, and not to ordinary names. Russell explicitly notes this difference in Appendix A of *PoM* § 476. If this feature made him vulnerable rather than Frege, it could only be because in 'On Denoting' he is arguing that the three-entity analysis is so excellent that it must be applied right across the board. In fact, of course, his later view is that it should not be applied to anything, and therefore engulfs Frege as well as his own more limited thesis.

So although there are these two respects in which the Russell of *PoM* differs from Frege, neither of them represents a weakness which he is later regretting, and it is not possible that 'On Denoting' refutes the former but not the latter.

Authors have not failed to invent features of Russell's earlier view, so that his argument can be seen as relevant to them, but not to Frege. Cassin, who believes that Russell did not even intend to attack Frege, holds that it was part of his earlier view (but not part of Frege's) that 'only terms could be denoted' but now Russell has come to realize that denoting concepts themselves must be denoted [2] p. 269. This is just wrong about the earlier view. A term, in *PoM*, is *anything* which can be the subject of a proposition, or an object of thought, or a logical subject (§ 47) and both things *and* concepts are terms (§ 48). Appendix B, § 483 criticizes Frege precisely for *his* difficulty over making concepts into logical subjects. And in § 476 Russell says:

> If one allows, as I do, that concepts can be objects and have proper names, it seems fairly evident that their proper names, as a rule, will indicate [i.e. refer to] them without having any distinct meaning [i.e. sense]; but the opposite view, though it leads to an endless regress, does not appear to be logically impossible.

The opposite view would have a higher order sense for any denoting phrase referring to a given concept. In Russell's terminology this would mean that when a definite description indicates a denoting concept, it itself has as its meaning another denoting concept which itself denotes (i.e. determines) the first. This is not a possibility he rejects.

And if he *had* rejected it in *PoM* it would have been a simple matter to change his mind without modifying his earlier view in any other way: argument which shows that he shouldn't reject it would leave the three-entity view quite intact.

Cassin (p. 270) also claims that there is another bad feature of the old view. Let us introduce the following abbreviations:.

C_1 = some denoting concept which denotes (i.e. determines) Aristotle

C_2 = some denoting concept which denotes (i.e. determines) C_1

According to Cassin it was part of Russell's old view that:

(A) any proposition containing C_2 as a constituent will be about Aristotle

We wanted it to be about C_1, but the concept C_1, since it determines Aristotle, makes the original attempt at denotation "fall through" to Aristotle anyway. It follows that it is impossible to refer to denoting concepts by means of expressions which have higher order denoting concepts as their sense, or in other words it is impossible to refer to them by descriptions, since such attempts fall through to the original object, in this case Aristotle. It would follow that denoting concepts of any level could at best be referred to by names, and this the later Russell might be finding objectionable. Searle [12] sees the argument in essentially the same way, only he regards (A) not so much as a feature of Russell's early view, but merely as a consequence of the conjunction of Frege's theory with its negation, which is all, according to him, that Russell succeeded in attacking.

There are two major problems with this view. Firstly (A) is not stated in *PoM*. Nor have we found any passage into which it could possibly be read. Secondly, had Russell surreptitiously held (A) he could hardly have failed to notice its effect on the regress of denoting concepts. Yet the passage we have quoted shows that at that time, when explicitly discussing Frege, he thought the regress quite possible. So to interpret Russell this way we have to suppose that he held a view which he never stated, in spite of the thoroughness of *PoM*, but was unfortunately incapable of drawing its most elementary consequences. It is but a short way from here to the theory of Jager [6], who holds that Russell was refuting a straw man especially developed to be refuted in 'On Denoting'. All these expositions offend against our cardinal principle of Russellian exegesis. Clearly none of them would appear plausible but for the difficulty of finding a sensible argument in 'On Denoting', against either Frege or *PoM*. So if we can present such a thing, they may be thankfully abandoned.

The same strictures do not apply to A. J. Ayer [1], who believes Russell to have a valid argument against Frege. Ayer correctly identifies the force of Russell's conclusion—that there is going to be a mystery about identifying senses and their relations to corresponding references —but not Russell's argument for that conclusion. In Ayer's exposition (p. 31) Russell is represented as demanding that 'the first line of Gray's Elegy' and 'the meaning of "the first line of Gray's Elegy" ' should have different meanings (senses) but the same denotation (reference). This

would of course be quite inadmissible for Frege: it is in fact just
the demand that sense and reference be identified. But Russell's only
reason for making that demand, on Ayer's account, amounts to com-
plaining that otherwise there will be a mystery about what the meaning,
or sense, is. This may be Russell's view, but then we still await an argu-
ment for it. We think instead that Russell had a powerful and precise
argument for supposing that Frege is doomed to ineradicable mystery,
and it is to the presentation of this that we now turn.

3. The argument.

We do not want to discuss tediously all of the terminological options
which paragraph (B) opens. We do not have to, since the point is to find
a sensible reading on which there is a good argument in the offing, not
myriads of readings on which there is not. Denoting phrases are of
course the expressions for which the theory has to work. Their meaning
Russell calls a denoting complex. This marks a departure from *PoM*
where he would have called it a denoting concept. It corresponds to the
Fregean *sense*. The thing denoted Russell calls the denotation. Denoting,
as before, is what we have called determining; it is the relation between
the sense and the reference (i.e., the denoting complex and the denotation).
Expressions, in this terminology, do not denote. Russell is using single
quotes, here at any rate, in order to give expressions which themselves
refer to senses, or denoting complexes. The relationship which he wants
us to consider at the end of (B) is that between sense and the reference
which it determines—i.e., determining.

His claim in (C) is that *if* an expression has a separate sense in
addition to its reference, *then*, fatally, there will be no guarantee that
there is a logical relation between the two. Or, in Russell's words,
'. . . we cannot succeed in *both* preserving the connexion of meaning
and denotation *and* preventing them from being one and the same'.
It is the business of (D)—(F) to show this, and the business of (G)—(H)
to show how the truth of this conditional would render Frege's theory
worthless. The point to be demonstrated is not that there is a difficulty
involved in the idea that a sense may itself be an object of reference, but
rather that there is a difficulty involved in specifying one in such a way
as to allow us to show that it performs a certain logical role. Let's see
how he does this.

The first important thing that he says about the determining relation
is that it cannot be 'merely linguistic through the phrase'. What he means
is that determining cannot be explained by Frege in terms of expressing
and referring. To understand this, we must remember a little of Frege's
theory. In Frege sense is a theoretical entity, and it is postulated that
denoting phrases come to refer to their normal senses when such phrases
are embedded in psychological contexts. This, as is well-known, is to

explain how it is possible that 'George IV wished to know whether Scott is the author of Waverley', and 'George IV wished to know whether Scott is Scott' should have different truth-values. Russell finds this explanation unsatisfactory. As pointed out above, this is not because of any pure difficulty about referring to senses. It is rather because in the absence of a theoretical definition of terms purporting to refer to senses, we cannot be sure what logical role sense plays. And Russell wants to show that there is simply no way to specify senses so as to be sure that they play the role Frege demands. To highlight this aspect of our interpretation, we would like to mention once again that in (C) the difficulty is over ensuring a certain connexion, and also point out that in (G) it is failure of explanation, residual mystery, inextricable tangle, that are said to menace Frege.

The Russell of *PoM* was quite content to take in the notion of a denoting concept (and likewise, Fregean sense) without requiring a careful theoretical introduction. Phrases referring to denoting concepts were simply taken as indefinable. But as the recalcitrance of the Russell paradox became apparent, he realized that inside even the most innocent theoretical constructions there might lurk fatal problems. Later we shall see that he also had good reason for a change of heart towards denoting concepts.

A specification of sense which does not guarantee its logical role would be to say *just* 'in order to speak of the sense of an expression "A" one may simply use the phrase "the sense of the expression 'A' " ' (Frege, p. 59. Of course, we do not dream of implying that Frege would have thought this to be sufficient by itself). If we postulate a sense by simply mentioning an expression and describing our entity as the sense of that expression, then we can explain the relation between sense and reference only by saying: well, it is the relation which holds between the sense of the term "Aristotle", and the man Aristotle, for example. This would be to make the relation 'merely linguistic through the phrase'. A forceful way of seeing why that will not do is to consider the pair:

    (1) Aristotle taught philosophy.
    (2) Jones believes that Aristotle taught philosophy.
and the pair:
    (3) Aristotle, the magnate, married Mrs. Kennedy.
    (4) Aristotle, the philosopher, wrote books.
In the latter case the relationship between the two things, referred to respectively by "Aristotle", is merely linguistic through the phrase. Now it is obvious that there must be a much closer connexion between the reference of 'Aristotle' in (1) and its reference in (2). It is not an accident of homonymy that the word recurs. Indeed, Russell is right in thinking that there must be a logical connexion between (1) and (2)— and therefore between the reference of 'Aristotle' in each—for they mate

together to yield the conclusion that Jones believes something true. So unless we are able to say something further about this connexion, the theory of sense will be quite unintelligible: there would be nothing but logically irritating punning to link the sentences (1) and (2). We have therefore an "adequacy-condition" on any theory of sense: we must be satisfied that on the theory this deduction is preserved. When, later, we query whether a proper introduction of the notion of sense can be given, it is this standard we have in mind. For now let us note that sense cannot be introduced or defined "linguistically". This is also likely to be the point of Russell's saying in (C) that it will be a difficulty if meaning can only be got at by means of, i.e., by mentioning, denoting phrases. So what we need is a definition of either 'the sense of "Aristotle" ' or some other phrase which we may suppose to refer to the sense of 'Aristotle'. The remainder of (D)—(F) shows grounds for supposing that this cannot be done.

After this preliminary, Russell turns, in (D), to head off a false move. Baffled at this objection to introducing sense via *mentioning* the expression, a Fregean might make the mistake of *using* it, and try to give the required definition of 'the sense of "Aristotle" '. He might define it as the sense of Aristotle. (Again, we do not mean to suggest that Frege would have made such an error.) (D) shows quite decisively what is wrong with this approach. Russell takes the case, favourable to the Fregean, where the thing referred to is a linguistic expression and might itself be supposed to have a sense. And he points out that even in that case, its sense—the sense of the reference—could not be the same as the sense of our original denoting phrase, which is the object we wish to specify. That is made clear by the Gray's Elegy example. 'Thus in order to get the meaning we want we must speak not of "the meaning of C" but of "the meaning of 'C' " . . .' This, of course, takes us back to the method of introduction which is merely linguistic.

It should dawn on us at this point that Russell has discovered a dilemma of sorts. We ask the Fregean to give a proper introduction to the notion of sense, and he can either mention expressions, or use them. If he mentions them, then the introduction fails, partly because it is scientifically improper just to postulate what he wants, but more seriously because the relation between sense and reference is left merely linguistic through the phrase. If on the other hand he uses the denoting phrase, then he refers, and you cannot introduce sense as the sense of the thing you refer to, for even if there is such a thing, that is not what we are after. Russell repeats the second horn at the beginning of (E): 'the moment we put the complex in a proposition, the proposition is about the denotation; and if we make a proposition in which the subject is "the meaning of C" then the subject is the meaning (if any) of the denotation, which was not intended'.

Thus all is plain sailing to the bottom of page 49. Russell has discovered his dilemma and given, in (D), a precise statement (with only minimal misuse of quotation marks) of the initial stage of the "use" horn of it. The "mention" horn is merely hinted at in (C), but as we have read it there is no solution there for Frege. Now at the end of (E) he reminds us of this horn: '. . . the meaning has denotation and is a complex, and there is not something other than the meaning, which can be called the complex, and be said to *have* both meaning and denotation'. The point of course is not that Russell has forgotten the Fregean view that denoting phrases will have both, but that he has already shown that leaving the matter there provides no explanation, or justification, of sense.

(F) is probably the most puzzling paragraph of all. But with a little care it becomes plain enough. The problem is still how we are going to specify this thing, the sense, meaning, or complex. Not, again, by using the denoting phrase in question, because then 'what is said is not true of the meaning, but only of the denotation, as when we say: The centre of mass of the solar system is a point'. The following two sentences then emphasize that when we refer to the sense, the expression we use must itself have its own sense which determines the sense we want to talk about. (Note: Russell is here using 'subject' not for the topic of discourse (the reference), but for the constituent of the proposition (Thought) which corresponds to it, i.e., the sense, meaning, complex.) Although these considerations apply to any expression purporting to refer to the sense we are after, Russell is here thinking primarily of the original defining expression of the sense (e.g., 'the sense of "Aristotle" '). And then comes the hammer: there is no way of identifying this secondary sense (e.g., the sense of 'the sense of "Aristotle" ') as a function of the original sense. For quite generally, there is no way of identifying a sense in terms of the corresponding reference: '. . . *there is no backward road from denotations to meanings, because every object can be denoted by an infinite number of denoting phrases*' (our italics).

Thus by the end of (F) Russell has powerfully sharpened the second horn of his dilemma. It is now not just that the phrase 'the sense of "Aristotle" ' will not do as our specification of the object which is, in fact, the sense of 'Aristotle'. This is what (D) pointed out. We are now threatened that *no* denoting phrase can specify the sense of a phrase by mentioning the reference of that phrase, and trying to identify the sense in terms of some fixed function of that reference. For since there is *no backward road*, this function would be quite unexplained. So there is a mystery as to how we define 'the sense of "Aristotle" ' or any other phrase which we may have reason to think refers to the sense of 'Aristotle'. But lacking such a definition, what guarantee is there that sense and reference are connected logically, as opposed to 'merely' linguistically?

We shall now state the case generally. Consider the following three expressions:

(E_1) 'Aristotle'

(E_2) 'the sense of "Aristotle" '

(E_3) 'the sense of E_2'

E_2 is a phrase whose reference is the sense of the name 'Aristotle'. If senses are to be properly introduced then there must be *some* denoting phrase, such as E_2, which we understand, and whose reference we can therefore grasp. The only alternative is that we need no such phrase because, somehow, we recognize senses (and their connexions) outright, without requiring a description or a definition. We discuss this further below, but obviously it is likely to leave senses in uninviting obscurity. However, the alternative that we recognize senses through our understanding of some phrase such as E_2 (although as the "mention" horn has shown, E_2 itself will not do) is open to us only if we are satisfied with that understanding. The trouble is that to understand E_2 is to grasp *its* sense, and the argument now repeats itself at the higher level. How are we satisfied that we grasp the sense of E_2? Either we recognize it straight off, or there is some denoting phrase, such as E_3, which refers to it, and which in turn we understand. But then, in turn, the lack of a backward road will entail that we need a separate story about that piece of understanding, and so a regress is generated. The regress can be stopped, but only if we pay the cost of saying that there is some level at which we do not need an understood description or a definition, but can rest content with an outright recognition of sense.[1]

Dummett ((5) p. 267) sees Russell's objection in a slightly different way. He writes that from Russell's 'extremely confused' criticism of Frege 'we can at least extract . . . a valid criticism of Frege's doctrine of indirect sense and reference'. But he also thinks that there is a 'simple emendation which can be made to the doctrine, which is in harmony with Frege's other views, and dispels the objection'. According to Dummett, Russell, noticing that Frege needs an indirect sense for a name to express in those contexts in which it refers to its own ordinary sense, objects that there is no way of telling what this thing is, because there is no backward road. Dummett's emendation is to identify indirect and ordinary sense, so that in such contexts a word expresses just the same sense as it ordinarily does: 'the sense of a word cannot vary from context to context, but is a property of the word itself, apart from any context' (p. 268).

This may be satisfactory as a doctrine about indirect sense, and therefore a reply to Russell as Dummett sees him. But it is vital to realize

[1] Tom Baldwin pointed out to us that on pp. 223–227 of 'Knowledge by Acquaintance and Knowledge by Description' Russell presents a closely related regress argument against Frege's view of identity statements.

that it has no force whatever against the difficulty as we have put it. The problem about the sense of a denoting phrase such as E_2 is not the same as the problem about indirect sense. They could be confused by supposing that the sense of a phrase such as E_2 might itself be the indirect sense of the name 'Aristotle'.[1] Dummett himself says that this is 'rather implausible' (p. 267). In fact it is not only that, but taken in conjunction with Dummett's identification of indirect and ordinary sense, would be totally lethal. For it would lead straight to the collapse of sense into reference. We need only consider the sentence 'Jones knows that Aristotle=Aristotle' and the sentence 'Jones knows that Aristotle=the sense of "Aristotle" '. E_2 in the context of the last sentence refers to its normal sense. If that is the indirect sense of 'Aristotle' and hence, following Dummett, to be identified with the ordinary sense of 'Aristotle', then the phrase E_2 in this context refers to the same as the name 'Aristotle' in this context. Each denoting phrase is referring to the ordinary sense of 'Aristotle'. Hence the two sentences must have the same truth-value, by the fundamental Fregean principle that if two sentences differ only by substitution of denoting phrases which, in the context of those sentences, refer to the same thing, then they cannot differ in truth-value. Hence since the first sentence is true, it follows that Jones knows that Aristotle is the sense of 'Aristotle', in which case Aristotle *is* the sense of 'Aristotle', and as Russell promised, we have failed to prevent them from being one and the same. The argument shows that the indirect sense of 'Aristotle', on Dummett's view of it, cannot be the same as the sense of 'the sense of "Aristotle" '. It follows that de-mystifying indirect reference is no help to Russell in his scepticism about how we are going to understand the purported defining expression of an ordinary Fregean sense. When in (*F*) and (*G*) Russell pursues the denoting complex 'C' which is to determine a sense, it is this which he is after, not the indirect sense required by Frege.

Of course, none of this is an argument that we could not refer to senses, once satisfied with their credentials. As we have stressed, the argument does not seem to be about reference. When Russell asks 'where are we to find the denoting complex "C" which is to denote C?' his plea is for a denoting phrase which we understand, the understanding of which we can explain, and which specifies the ordinary sense. A phrase such as 'the sense of "Aristotle" ' fails these conditions; so does a phrase which uses the word rather than mentioning it. If such a phrase is apparently intelligible to us ('the sense which each of us grasps when he thinks of Aristotle', for instance), it nonetheless fails as a definition by presupposing that we already know what we are talking about. It does not tell us what we are talking about, because the lack of a backward

road means that we would have no idea of how the word 'Aristotle' would function in such a description.

We did notice the option of supposing that senses are named out-right, presumably with an attempted explanation of what is so named, rather than introduced with a definite description referring to them via their defining properties. It is worth noting at this point that (although Russell did not know this at the time) the *PoM* theory of denoting makes it impossible to directly name a sense or denoting concept. This is because (i) when a name is used in a sentence, the thing named is a constituent of the proposition expressed by the sentence, and (ii) if a denoting concept is a constituent of a proposition, then the proposition is about the object denoted and not about the denoting concept.[1] If we suppose that he had come to see this by the time he wrote 'On Denoting', we have an explanation for the fact that he is now insisting that senses be introduced by means of definite descriptions. (Notice that although this *is* a problem as to how one refers to senses, if there are such things at all, the obvious solution is not to attack Frege, but rather to insist that his three-entity view applies to *all* referring expressions.)

The heart of the matter, then, is whether the lack of explicit definition can be tolerated. Will explanations and indications which fall short of this still suffice to give us confidence in the notion? In a different passage (p. 227) Dummett writes that in saying what the reference of a word is we *show* what its sense is, and conceding that we cannot directly state what the sense of an expression is he offers us the con-solation that we need only sufficient grasp of the notion to say what it is that someone can do when he has grasped a sense. But this consolation is empty. Frege's theory demands that we do refer to the senses of denoting phrases. We refer to them, for instance, whenever we put names in indirect contexts: our understanding of any of these is as frail as our understanding of precisely what senses are. The truth-conditions of all ascriptions of belief, thought, knowledge, are no more apparent than the properties of senses, nor are the relations between these things and the ordinary world of objects any more luminous than the relation between a sense and the object which it determines. Thus consider the common explanation in terms of a "way of finding a reference", a "mode of presentation of a reference", or a "criterion for identifying an object as being the bearer of the name". Do we understand these things and their relation to ordinary objects of talk? What is the logical connexion between Aristotle and a mode or way of finding him? How is it that a sentence such as (1) combines in a deduction with a sentence such as (2) about this very different thing? (Perhaps we should pause and remember the fate of many attempts to illuminate the logical relationship between criteria and things for which they are the criteria.) It is no part of our

[1] Again, David Kaplan showed us this.

argument that these questions are unanswerable, but it is our contention that Russell's argument is strong in proportion as their answers are obscure. Logic may be ineffable but it is a pity to build its ineffability into the very bottom of our theory about the relation between thoughts and things. Perhaps a theorist can plead some licence for intuitive notions, and not everything can be defined—but is it too exacting a critic who complains when the undefined construct has indefinable but firm and clear logical relations?

Russell's solution is to give sentences containing definite descriptions a quantifier-plus-propositional-function interpretation, thus leaving the object referred to no place in either our understanding of them or our account of what makes them true. It has been suggested by Kaplan that since this is available we may simply identify the Russellian propositional function with Fregean sense, thus meeting any Russellian qualms about the ontological and logical status of senses, although not satisfying his distrust of a mixed theory and the logical (sense to reference) connexions it must have. Whether or not such a move will in the end prove satisfactory (will it provide a sense for phrases such as 'the sense of "Aristotle" ' which Russell does not contextually define, for instance?), we must still keep it in mind that prior to his development of the theory of descriptions it could not have been made at all. Hence the objections to Frege could not have been met. It is hard to take away from Russell the credit for just the respect in which his theory was shatteringly new. It may be that for some or many denoting phrases Frege's theory can be resurrected, but we are certain that in 1905 Russell knew of difficulties for it which, in the intervening seventy years, have seldom been noticed, and never laid to rest.

© SIMON BLACKBURN AND ALAN CODE 1978

Pembroke College, Oxford and University of California—Berkeley

[1] A. J. Ayer, *Russell and Moore: The Analytical Heritage*, Macmillan (London, 1971), 30–32.
[2] R. J. Butler, 'The Scaffolding of Russell's Theory of Descriptions', *Philosophical Review*, LXIII (1954), 350–64.
[3] Chrystine E. Cassin, 'Russell's Discussion of Meaning and Denotation: A Re-examination', *Essays on Bertrand Russell*, ed. E. D. Klemke, University of Illinois Press (1970), 256–72.
[4] Alonzo Church, 'Carnap's Introduction to Semantics', *Philosophical Review*, LII (1943), 298–304.
[5] M. Dummett, *Frege: Philosophy of Language*, Duckworth (London, 1973).
[6] Gottlob Frege, 'On Sense and Reference', *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. T. Geach and M. Black, Basil Blackwell (Oxford, 1952), 56–78.
[7] P. T. Geach, 'Russell on Meaning and Denoting', *Analysis*, XIX (1959), 69–72.
[8] R. Jager, Russell's Denoting Complex', *Analysis*, XX (1959), 53–62.
[9] Bertrand Russell, 'On Denoting', *Logic and Knowledge*, ed. R. C. Marsh, George Allen & Unwin Ltd. (London, 1956), 41–56; reprinted from *Mind*, n.s., XIV (1905), 479–93.
[10] Bertrand Russell, *The Principles of Mathematics*, 2nd ed., George Allen & Unwin Ltd. (London, 1937; first ed. 1903).
[11] Bertrand Russell, 'Knowledge by Acquaintance and Knowledge by Description', *Mysticism and Logic*, George Allen & Unwin Ltd. (London 1917) pp. 223–227.
[12] J. R. Searle, 'Russell's Objections to Frege's Theory of Sense and Reference', *Analysis*, XVIII (1958), 137–43.

FUGITIVE PROPOSITIONS AGAIN

By RAINER BÄUERLE

IN his article 'Fugitive Propositions' (ANALYSIS 10.1, 1949, pp. 21–3), Austin Duncan-Jones raised the problem that, given the usual rendering of the past tense statement 'Brutus killed Caesar' as

   (a) 'For some t, Brutus kills Caesar at t, and t is before now',

it is impossible for any two historians to agree or disagree with each other: if they use (a) on different occasions, the proposition entertained will be different.

In a rejoinder, Patrick Nowell Smith ('Fugitive Propositions', ANALYSIS 10.5, 1950, p. 100–3) remarks that there is nothing mysterious about sentences that have a different meaning each time they are used, because such sentences contain an implicit reference to the context of use. But although this helps us towards a better formulation of the problem (Duncan-Jones took 'now' to be the proper name of a moment), it certainly does not solve it.

A solution was attempted, however, by L. Jonathan Cohen ('Tense Usage and Propositions', ANALYSIS 11.4, 1951, pp. 80–7). He offered the following analysis for the sentence in question:

> 'For some t, Brutus kills Caesar at t, t is the point or period of time under reference, and the time under reference is prior to the time at which any member of the following class of tokens is uttered, written or thought, viz. the class consisting of this and all other past-referring tokens which would normally be thought to be in direct agreement with it.'

Now I do not want to discuss what exactly 'would normally be thought' is intended to contribute to the argument. The main point, it seems to me, is that Cohen has made into a premiss of his argument what he should have explained: namely, what it means for two tokens to be in direct agreement with each other. Although he correctly remarks that the intuitive recognition of a class of such tokens is a prerequisite of any analysis, surely there is nothing wrong with trying to explain one's intuitions. And that was indeed what Duncan-Jones had in mind when he observed that, although the propositions entertained are different, we intuitively feel that two historians want to dispute one and the same thing, or, as he put it, entertain the same proposition.

I shall now argue that at the heart of the matter lies a mistaken analysis of tenses. It can easily be shown that

   (b) For some t, John comes at t, and t is before now

cannot be a sufficient rendering of

   (c) John came,

for the simple reason that in ordinary discourse, even if I use (c) today to make a true statement, I may use it again some time in the future and make a false statement. This cannot be done with (b): if there exists a t before now, such that John comes at t, it cannot be wrong at any future time t^1 that there is a t before t^1 such that John comes at t. That is to say, if (b) is true now, it will henceforth always be true.

. Now I do not want to argue that we never use (c) in the sense of (b), but only that (b) seems to be a rather special use, depending on some contextual information like the following:

Have you *ever* been to London?—Yes, I have (been to London).

The answer to this question, if true now, will of course remain the true answer for any future utterances of this question, when addressed to me.

In order to see how the analysis of tenses has to be improved, we must also take into consideration that it is not only when we use one and the same sentence at different times that we entertain different propositions. Even if by any chance my wife and I were to utter (c) simultaneously, though speaking to different people, I may have done so because asked for an explanation for my absence at the colloquium on the previous evening, whereas my wife could have answered the question why we failed to join our friends at the cinema a week ago. In such a context, what I am trying to communicate is that John came the evening before, whereas my wife asserts that John came the very time we wanted to meet our friends at the cinema last week. In other words: it is not only the time of utterance that is relevant to an interpretation of the utterance of (c). We cannot fully understand what is asserted with an utterance of (c) when we only know the time of utterance, we must also know what time the speaker is referring to.

. The same line of argument can be applied if we use the sentence at different times. But we should now be aware that the real problem is not to account for the moving 'now', but to explain how we come to know whether the sentence, when used on different occasions, refers to the same event or to different events. This is what Cohen failed to do. And indeed, the original sentence 'Brutus killed Caesar' is ill suited to start an analysis with: our knowledge tells us which event the speaker must be referring to, because we know of only one such event. It is harmless everyday sentences like (c) that most clearly show us the importance of two time-indices for an understanding of tense.

The time referred to is either supplied by the context of the utterance, or else made explicit in the sentence by means of a temporal adverb. In the latter case tense is used anaphorically. The function of tense is to specify some subinterval of the time interval referred to as the interval during which the event or events referred to occurred. In the case of the past tense, the subinterval is that part of the time referred to that is

earlier than the time of utterance: the past part of 'today' in 'John came today', the past part of 'yesterday' (i.e. all of yesterday) in 'John came yesterday'.

On the basis of such an analysis we can be a bit more precise as to what it means for utterances to be in direct agreement with each other. Utterances can be said to be in direct agreement if they refer to the happening of some event or events during a particular reference-time. And we need no longer be puzzled how two historians can possibly agree or disagree with each other, even if they use the same sentence at different times: they also agree or disagree on the happening of some event or events during a particular reference-time.

Universität Konstanz   © RAINER BÄUERLE 1978

SIMILARITY AND COUNTERFACTUALS

By EUGENE SCHLÖSSBERGER

DAVID Lewis in *Counterfactuals* (Blackwell and Harvard University Press, 1973) claims that the notion of comparative similarity provides a 'foundation for the clarification of counterfactuals'. In fact, however, comparative similarity and counterfactual truth fail to 'coordinate' in several important ways. In this paper I shall consider one such way.

Lewis's approach is to place all (accessible) possible worlds in a system of 'spheres', or centred subsets of possible worlds. Thus sphere S_2 will include all the worlds in sphere S_1, but not conversely. These spheres of possible worlds are ordered by degree of similarity to the actual world. Hence the actual world is at the centre of the system, sphere S_1 will consist of those worlds most closely resembling the actual world, sphere S_2 will include all the worlds in S_1 plus those worlds which are but slightly less similar to the actual world, etc. According to Lewis, a counterfactual p is true iff there is a sphere S_n such that (a) S_n contains at least one antecedent satisfying world (world in which the antecedent of p is true), and (b) all antecedent satisfying worlds in S_n are also consequent satisfying worlds.

Lewis's analysis, then, instructs us to seek out those antecedent satisfying worlds which most resemble the actual world, and to ignore all other possible worlds. However, some counterfactuals must apply to a range of possible cases in order to be true, not only to those cases most similar to the actual world.

For example, imagine a dart-board rigged to a bomb in the following way: if a dart strikes the board at point p, the bomb will not go off. If a dart strikes the board anywhere else within a small circle A (of which p is the centre), the bomb will go off, and if a dart should strike the board anywhere outside of A the bomb will not detonate. In the actual world I throw a dart at the board which lands at p. Consider the following proposition:

> (1) If the dart hadn't struck the board at p, the bomb would have gone off.

Now surely those worlds in which the dart strikes the board within A are (*ceteris paribus*) more similar to the actual world than those in which the dart strikes outside of A (since a smaller displacement of the dart is involved). Thus there will be a sphere S_n in which all of the worlds in which the dart does not strike at p are worlds in which the dart lands inside A (and hence worlds in which the bomb detonates). So on Lewis's account (1) is true. But (1) shouldn't be true. For although

> (2) If the dart hadn't struck the board at p, the bomb might have gone off

is true, (1) is true only if the bomb would detonate as a result of any plausible landing of the dart. And although 'plausible' is somewhat hard to define in this context, surely the world in which the dart lands at a spot just outside of A is a defeating counterexample to (1). Such a world, however, is less similar to the actual world than many worlds in which the dart lands within A, and so is excluded from consideration by Lewis's analysis.

There is another difficulty for the Lewis analysis lurking here. One might say that any world in which the bomb goes off is more divergent from the actual world than any world in which the bomb fails to explode, regardless of where the dart lands, since the detonation of a bomb is a matter of much more moment than the point of impact of a dart. Presumably, the world in which the dart misses the dartboard and the bomb doesn't detonate is more similar to the actual world than any world in which the bomb detonates. So it seems to be true on Lewis' account that even if the dart hadn't struck the board at p, the bomb would still not have gone off. Notice that Lewis cannot avoid this by claiming that only events and states of affairs up to and including the moment of impact (call it t_n) may be used to compute similarity to the actual world. For a world H in which the dart lands on a spot q within A but a flash flood shorts the bomb mechanism before it explodes is just as similar to the actual world prior to $t_{n+\epsilon}$ as is the world J in which the dart lands at q and the bomb explodes (indeed, H and J are indistinguishable prior to $t_{n+\epsilon}$).

Lewis could, of course, claim that for the purposes of (1) any world in which the dart strikes the board at all is, *ceteris paribus*, equally similar

to the actual world. Such a move, however, seems to me to be tantamount to giving up the analysis in terms of similarity altogether. For similarity is otiose here unless either the relevant kinds of similarity and their appropriate weights are clearly demarcated in advance, or one can trust one's ordinary intuitions about similarity. If similarity is to do any work, either (*a*) Lewis must delineate a set S of kinds of similarity $s_1 \ldots s_n$, and a set C of types of counterfactuals $c_1 \ldots c_n$, and provide a mapping of members of S to each member of C (i.e., by assigning to each $c_n \epsilon C$ a unique weighted subset of S), or (*b*) one must be allowed to depend upon one's ordinary intuitions about which of two worlds, H and K, is more similar to a given world J. Otherwise one might as well drop reference to similarity and rest content with an unanalysed ordering relation instead.

Thus whether or not a possible worlds approach to counterfactuals is feasible, comparative similarity does not seem to provide the right ordering principle.

© EUGENE SCHLOSSBERGER 1978

TOO MUCH OF A GOOD THING:
A PROBLEM IN DEONTIC LOGIC

By ALAN McMICHAEL

AMONG the items of ordinary discourse, we find expressions of moral obligation, such as 'You should be kind to animals' and 'You ought to help needy people'. We also find expressions of conditional obligation, such as 'If you make a promise, then you should keep it' and 'If you offend someone without cause, then you ought to apologize'. Deontic logic is the study of inferential connections between these and related expressions. Those who have studied the subject know that fallacies lurk everywhere. For instance, if 'O' is an operator of un-conditional obligation, then neither the form 'O($\varphi \supset \psi$)' nor the form '$\varphi \supset O(\psi)$' is suitable for expressing conditional obligations. And if '$\square \rightarrow$' is a connective for conditional obligation, then '$\varphi \square \rightarrow \psi$' may be true when '($\varphi$ & χ)$\square \rightarrow \psi$' is false. Much recent work in deontic logic has been devoted to giving precise truth-conditions for deontic statements, truth-conditions which are supposed to explain the fallacies.

Attempts have been made to formulate truth-conditions using two working notions, that of a proposition (or sentence) being true at a possible world, and that of one possible world being better than another. An elegant and fully sophisticated treatment of this sort has been given by David Lewis in his book *Counterfactuals*.[1] Lewis employs his principles quite successfully to account for the various deontic fallacies. He correctly observes that these fallacies are analogous to similar fallacies in the logic of counterfactuals. Nevertheless, it seems to me that his truth-conditions are flawed in a way that calls into question the whole approach. And this approach, the attempt to give truth-conditions in terms of an ordering of possible worlds, is by no means peculiar to Lewis's work.

Lewis says, roughly, that a conditional obligation '$\varphi \square \rightarrow \psi$' (read: Given that φ, it ought to be that ψ) is true from the standpoint of a world i if and only if either (1) φ is not true at any world evaluable from i, or (2) for some world w evaluable from i, φ is true at w and for every world w' evaluable from i, if w' is at least as good as w from the standpoint of i, then '$\varphi \supset \psi$' is true at w' (p. 100). In what follows, I think it is safe to assume that all possible worlds are evaluable and that standpoint does not matter.[2] By making these assumptions, we arrive at a simplified principle: '$\varphi \square \rightarrow \psi$' is true if and only if either (1) φ is impossible, or (2) for some world w, φ is true at w and for every world w', if w' is at least as

---

[1] David Lewis: *Counterfactuals* (Blackwell and Harvard University Press, 1973).
[2] In Lewis's terminology, these are the assumptions of *Universality* and *Absoluteness* (p. 99).

good as w, '$\varphi \supset \psi$' is true at w'. This notion of "conditional obligation" seems too strong to be useful.

I am inclined to believe that there is a good, perhaps happiness is such a good, which may exist in amounts of any size.[1] But if I am right, then all of the usual examples of conditional obligation fail to satisfy Lewis's truth-condition. Consider Lewis's example: given that Jesse robs the bank, it ought to be that he confesses and returns the loot. (p. 102). Select any world w, however good, at which Jesse robs the bank (φ). Then there will be a world w' which is better than w, but in which Jesse robs the bank without confessing and returning the loot ($\sim(\varphi \supset \psi)$). This is so because in some such world w', there will be enough extra good to counterbalance the absence of a confession and surpass the goods of w. In other words, since there is a good which may exist in amounts of any size, the lack of a confession puts no bound on the goodness of the worlds in which Jesse robs the bank. Consequently, Lewis's truth-condition yields the undesirable result that it is *not* true that given that Jesse robs the bank, it ought to be that he confesses and returns the loot. Virtually every other example of conditional obligation one can think of disappears as well.

Lewis gives corresponding truth-conditions for conditional permission, unconditional obligation, and unconditional permission. Each has a corresponding difficulty. The conditions of obligation are too strong, the conditions of permission are too weak. For example, Lewis says that an unconditional permission '$\Diamond \varphi$' (read: It is permissible that φ) is true just in case for any world w, there is a world w' which is at least as good as w and at which φ is true (p. 101). If there is a good which may exist in amounts of any size, then very horrible things turn out to be unconditionally permissible. Select any world w, however good. There is a world w' which is better than w, but in which Jesse gratuitously inflicts extreme pain on many kindly scholars. To be sure, there would have to be counterbalancing goods in w', but I see nothing to prevent their appearance.

If there is a good which may exist in amounts of any size, then Lewis's principles yield strange and unacceptable results: too little is obligatory, too much permissible. If there is no such good, there is at least considerable doubt about that fact. This doubt does not deter anyone from using the ordinary deontic notions. I conclude that those notions are not so sensitive as the ones for which Lewis has given truth-conditions. We are left without sound tests of the truth of ordinary deontic statements, without sound tests of the validity of deontic inferences.

University of Massachusetts at Amherst   © ALAN McMICHAEL 1978

[1] This leads me to reject the *Limit Assumption* (p. 98). For many propositions φ, there are no *best* worlds in which φ is true. Take any world in which φ is true. Can't we imagine adding some good to it without affecting the truth of φ?

REPLY TO MCMICHAEL[1]

By David Lewis

DEONTIC conditionals, whether those of ordinary discourse or the simplified versions invented by intensional logicians, are ethically neutral. You can apply them to state any ethical doctrine you please. The results will be only as acceptable as the doctrines that went into them.

Radical utilitarianism, stark and unqualified, is not a commonsensical view. Agreement with our ordinary ethical thought is not its strong point. It is no easy thing to accept the strange doctrine that nothing at all matters to what ought to be the case except the total balance of good and evil[2]—that any sort or amount of evil can be neutralized, as if it had never been, by enough countervailing good—and that the balancing evil and good may be entirely unrelated, as when the harm I do to you is cancelled out by the kindness of one Martian to another.

Accept this strange doctrine, and what should follow? Exactly the strange consequences that McMichael complains of! Never mind the semantics of deontic conditionals. If you really think that only the total matters, then surely you ought also to think that little is obligatory (there are always alternative ways to reach a high total) and that much is permissible (no evil is so bad that it cannot be neutralized). It is not in the radical utilitarian spirit to believe in outright ethical requirements or prohibitions.

Order the worlds on radically utilitarian principles; then apply the semantics for deontic conditionals that I gave[3] in *Counterfactuals*; and the results are as McMichael says they are. Most of us would indeed find these results strange and unacceptable, but the radical utilitarian should find them much to his liking. The semantic analysis tells us what is true (at a world) under an ordering. It modestly declines to choose the

---

[1] Alan McMichael, 'Too Much of a Good Thing: A Problem in Deontic Logic', ANALYSIS, this issue. McMichael there criticizes Section 5.1 of my *Counterfactuals* (Blackwell, Oxford, 1973).

[2] Since we are discussing my treatment in *Counterfactuals*, our topic is what ought to be, not what someone in particular ought to do. See my footnote on page 100. But parallel questions would arise for the deontic logic of personal obligations.

[3] I did give the semantic analysis under discussion—but I did *not* give it as an exact analysis of any 'items of ordinary discourse'. Rather I meant it as a stipulation of truth conditions for deontic conditionals similar to those already studied by some deontic logicians. These have their interest partly because of their resemblance to the deontic conditionals of ordinary discourse. But I fully agree with McMichael (though for different reasons) that the resemblance is far from perfect. Section 5.1 of *Counterfactuals* is not an essay in ordinary language philosophy. As I stated at the outset (page 96), it is a study of the formal analogy between counterfactuals and variably strict conditionals in deontic logic. I did say that those conditionals 'may be read as' certain constructions of ordinary English. (One might likewise say that the standard existential quantifier 'may be read as' some English construction, though aware of differences between the two.) Surely to say that is to claim nothing more than an approximate likeness of meaning. Since the differences I believe in between my deontic conditionals and those of ordinary language are irrelevant to the difference that McMichael believes in and I do not, I have here ignored them.

proper ordering.[1] That is work for a moralist, not a semanticist. If what
turns out to be true under a utilitarian ordering is what is true according
to radical utilitarianism, not what is true according to our ordinary
opinions, that is just as it should be.

Other orderings, other results. For instance, a simplistic non-
utilitarian might fancy an ordering on which the better of any two worlds
is the one with fewer sins. (It is up to him to tell us how he divides the
totality of sin into distinct units.) Under this ordering and my semantics,
much is obligatory and little is permissible. Perhaps some of the worlds
where Jesse robs the bank have sixteen sins, none have fewer, and some
have more. Then what is obligatory, given that Jesse robs the bank,
is that there be no seventeenth sin. No course of action with any extra
sin is (even conditionally) permissible, no matter how much counter-
balancing good there may be. McMichael's argument cannot be made in
this case. The only relevant good, sinlessness, is not 'a good which may
exist in amounts of any size'.[2]

What is true under a utilitarian ordering or a sin-counting ordering
(according to my semantics) ought not to be expected to agree with our
ordinary opinions. Ordinary moralists are neither radical utilitarians nor
sin-counters. It would be better to ask: is there *any* ordering (more
complicated than those yet considered, no doubt) such that what is
true under that ordering agrees with our ordinary moral opinions?

But even that better question is not good. Is there really any definite
body of "ordinary moral opinions" to agree with? I think not. We
disagree, we waver, we are confused. Few of us singly, still less all of us
together, have achieved a stable equilibrium between our utilitarian
and our sin-counting inclinations.

Princeton University   © DAVID LEWIS 1978

---

[1] See page 96.

[2] We might also consider an ordering in which the world with fewer sins is better, but in
which ties between worlds with equally many sins are broken on utilitarian considerations.
Even though we now have a relevant good which may exist in amounts of any size, it remains
true that much is obligatory and little is permissible. Avoidance of extra sin is obligatory,
given that Jesse has robbed the bank, because no amount of good can outweigh an extra sin.

HEIR OF FRANKENSTEIN
OR
JUSTIFIED TRUE BELIEF AND PERSONAL IDENTITY

By JONATHAN HARRISON

IT has been commonly, though not universally, held by philosophers that in order for A to know some proposition p, three conditions must be satisfied. (i) p must be true. (ii) A must believe p, and perhaps be quite certain that p is true. (iii) A must have either good or conclusive reasons for thinking that p.

The third of these conditions leads to an infinite regress. It is not enough that there should be some propositions, say q, r and s, which imply p. These propositions must be true propositions, or at least some of them must be true, and A must know them to be true, or, at least, know that enough of them are true to give him good or conclusive reasons for thinking that p is true. However, a definition of knowledge which stipulated that A could not know p unless he also knew some propositions, q, r and s, which gave him some reason for thinking that p, or provided him with conclusive reasons for p, would be circular.

Not only is the above definition of knowledge circular, it is fairly obvious that we cannot believe all we believe because we infer it from something else. At some point or other we must stop without having any reasons at all for what we claim to know, at any rate, if by 'reason for p' we mean 'proposition which supports p'.

There is no difficulty, of course, in finding propositions which we do know, without inferring them from anything else. We may know that we are in pain, that we are trying to solve a philosophical problem, that something in the centre of our visual field *looks* red, or that we were in too much of a hurry to eat our breakfast, without our inferring these things from anything at all, and so without having good or conclusive reasons for thinking them to be true.

This objection will not apply to the view that we must be justified in believing something if we can properly be described as knowing it. We can clearly be justified in believing that we are trying to solve a philosophical problem or that we did not have time to eat our breakfast, even though we had no reasons for believing these things, in that there were no other propositions from which we inferred them. I do not here wish to argue the matter any further, however, for there is a piece of science fiction which might well demonstrate not only that we can know some things to be true without having good or conclusive reasons for believing them, but that we can know some things to be true without being justified in believing them either.[1]

---

[1] The first part of the following story was suggested by an argument in 'An Analysis of Factual Knowledge', by Peter Unger, in *The Journal of Philosophy*, Vol. LXIV, 1967.

In the year 2977 A.D. a remote descendent of Frankenstein at last discovered a way of making complete human beings. It is true that he had the benefit of hundreds of years of the unsuccessful, but nevertheless enlightening, efforts made by numerous teams of physiologists and surgeons, supported by simply enormous grants from governments trying to overcome the problems presented by internationally declining birthrates. Some of these teams, indeed, had even succeeded in making frogs and mice, and one moronic anthropoid which behaved so badly that the team which produced it went into liquidation and destroyed its apparatus. Dr. Frankenstein, however, with only a meagre grant from the relatively impoverished British State Government, succeeded in making not a monster, but a perfect adult human being, indistinguishable in appearance, behaviour, and in all physiologically testable ways from any normal human who, unlike Frankenstein's creation, had a father and a mother. (Clearly there was no point in making human beings who were not adult, when it was just as easy to make a fully grown man as to make a baby, and the latter saved the time and expense of educating and suitably indoctrinating him.)

Frankenstein, who did not believe that the mind was necessarily a *tabula rasa*, saw that it would save him or his sponsors a great deal of trouble if his creation were brought into the world with all the necessary skills and information which a young man of his biological—though not, of course, chronological—age would normally have. So he planted in the brain of his creation—whom he decided to call Smith, after a colleague, for he was a man of little imagination—structures of neurons which, when the final life-giving current was switched on, would give Smith all this knowledge, without other people being put to the trouble of teaching it, or him himself of learning it. Hence, from the very first moment of his life he was possessed of such useful pieces of information as that the Battle of Hastings—still not completely forgotten—was fought in 1066, that the earth was round, the sun 93,000,000 miles away from it, and that the square on the hypotenuse of a right angled triangle equalled the sum of the squares on the other two sides. He was also given all the information he needed about the behaviour and social customs of the men who were to be his contemporaries.

Frankenstein, who was a kind man, realised that the trauma of creation might well be distressing for Smith, and so he implanted in Smith's brain all the modifications which would have been traces of a normal upbringing, if Smith had had one. Hence Smith not only had a perfectly adequate knowledge of History, Geography, Literature, Mathematics, Physics, Astronomy—much more than would be possessed by a professor of these subjects living today—but he also had seeming recollections of a father and mother, of companions, school days, and, in particular, of the lessons in which he thought—wrongly, of course—he had

acquired the knowledge with which Dr. Frankenstein had so benevo-
lently endowed him. Frankenstein's lack of imagination, however,
caused him to refrain from inventing a past for Smith. Instead, he mod-
elled Smith on the deceased son of the aforesaid colleague and his wife,
who connived at Frankenstein's enterprise; indeed, having only recently
lost their own son, they were delighted to accept Smith as a substitute.
They considered that the filial devotion—more common then than it is
now—of a man exactly resembling their child and who, albeit mistakenly,
regarded them as his parents was, at their age, a much better bet than
trying to replace their son in the normal way by someone who was really
their own. (In those days, fertility, which was in any case low, varied
almost exactly in inverse ratio to intelligence, and the intelligence of
Smith's foster parents was high.)

Smith, I am glad to be able to say, lived a happy and successful life,
and was a source of great delight and consolation to his foster parents,
and of pride to Dr. Frankenstein. Smith's success, indeed, did much to
soften the blow which fell in the shape of a refusal by the world govern-
ment to allow Frankenstein a licence to produce men any more intelligent
than was indicated by an I.Q. of 90 (a score relatively much lower then
than now) for certain routine tasks such as child minding and nursing,
for which computers had been discovered to be unsuitable. He did how-
ever, present a problem to the philosophers of his day—philosophy,
unlike every other academic discipline, had progressed little, if at all,
from the state it is in at the moment—who positively disapproved of
believing anything which one had not either proved oneself or which,
alternatively, was not firmly based in some way or other on one's
personal experience, and who certainly would not have counted beliefs
not so based as knowledge. Smith, though he could prove, when asked,
all the mathematical theorems he claimed to know, knew a large amount
of mathematics which he had never actually proved, for proving it all
would have taken months, if not years, and he had not been alive for
nearly that long. And though he seemed to know a large amount of
history, he claimed he knew these things because he had been taught
them by teachers whom he could not have had, or to have read them in
books to which he never could have had access. Though he expected
cats to purr and dogs to bark—in fact, he claimed to know that they did
—when people enquired how he knew these things, he would reply that
he had observed a large number of cats, all of whom purred but did not
bark, and a large number of dogs, all of whom barked but did not purr.
To the best of Dr. Frankenstein's knowledge, however, he had never
seen a cat or a dog in his life. Smith, however, delighted a small minority
of philosophers, who thought that it was unnecessary for a man to
justify his beliefs, in order for him properly to be said to know the things
he believed, provided that there was a causal connection between the

facts he claimed to know and his beliefs, such that he would not have believed them had they not been facts. They could point out that there was such a causal connection in Smith's case for, to take just one example, if the Battle of Hastings had not been fought in 1066, Dr. Frankenstein would not have believed that it was, and, if he had not believed that it was, he would not have caused Smith to believe this too.

The traditionalist philosophers did evolve a reply to their opponents —in fact, two replies—though it took them many years to hit on them. The first was that, since Dr. Frankenstein, though a little dull, was absolutely reliable, Smith *was* justified in believing the things Dr. Frankenstein had made him believe. He did not, of course, know that he was justified in believing them, because he did not know that Dr. Frankenstein had implanted these beliefs, or know that he was the creation of Dr. Frankenstein at all, but nevertheless he was justified in believing them, and so did know them, although he did not know that he knew them. The second was that, since there was a causal connection between what Smith's foster parents' deceased child, Smith minor, had learned, and what Smith knew, Smith simply was Smith minor in another incarnation. Dr. Frankenstein had thought it kindest both to Smith's foster parents and to Smith not to give him any information except that which he knew had been possessed by Smith's archetype, Smith minor; hence if Smith minor had not known it, Dr. Frankenstein would not have believed that he knew it, and if Dr. Frankenstein had not believed that Smith minor knew it, he would not have caused Smith to believe it. So Smith not only resembled Smith minor physically, and had the same aptitudes, personality, character traits and information as Smith minor, there was also a causal connection, *via* the skill of Dr. Frankenstein, between Smith minor's acquiring the information, the personality and the character traits and Smith's having the information and the personality and the character traits. Hence Smith, the traditionalists argued, *had* proved or learned from experience the facts he claimed to know, for Smith minor had proved or learned them, and Smith minor just was Smith.

Smith's foster mother (or mother, depending upon which school of thought you happened to belong to) was so attached to her adopted son (or son) that she took a great interest in the controversy, and after a rigorously careful examination of the arguments on each side of the question, concluded that the traditionalists about knowledge were right. Smith, she decided, must actually be her son, and she herself had imparted a great deal of the information, and almost all the moral precepts, which the heretical philosophers thought he knew without being justified in believing them. Since the embargo on Dr. Frankenstein's creating another man with an I.Q. of over 90 prevented him from making another Smith, she was happily saved the distress of having to decide whether two exactly similar men were each identical with the boy she had so

successfully nurtured, and were each her only child. Similar problems were, it is true, caused by Dr. Frankenstein's uninventiveness in producing large quantities of exactly similar nurses and child minders, all modelled on the daughter of the woman who cleaned for him, but that is another story, and Smith's mother herself took no interest in them.

University of Nottingham © JONATHAN HARRISON 1978

THE ADDRESSING FUNCTION OF 'I'

By D. S. CLARKE, JR.

ANSCOMBE has argued recently[1] that the pronoun 'I' is not a referring expression, and hence there is no need to postulate a 'self' as its referent. The view that 'I' does refer (one shared by Descartes and Strawson) rests on a grammatical illusion, the illusion that what occupies the place of a grammatical subject must be a logical subject. She attempts to establish this conclusion by showing the difference in function between the third person pronoun and the first person. For the latter, she contends, it is impossible to substitute either proper names or demonstrative phrases. For me to say 'I am sitting' is not to say 'Clarke is sitting' or 'This person is sitting'. Since it is not proxy for either a name or a demonstrative phrase, it follows that 'I' is not used as a referring expression.

Anscombe's conclusion is, I think, correct. But it rests on faulty premises. In fact, there is a form of demonstrative phrase functioning as does 'I' and contexts in which proper names can be substituted for the pronoun. To specify both is to specify the function of the pronoun and contrast it with that of a referring expression.

First, a brief review of Anscombe's main argument. It appeals to the fact that for singular referring expressions there is both the possibility of the speaker not knowing whom or what he is referring to and that the purported referent does not in fact exist. This is obvious for definite descriptions, but occurs also for names and demonstratives. If I say 'John Smith inherits five thousand dollars' I may not know who John Smith is or even that he exists. Even if John Smith says while reading a will 'John Smith inherits five thousand dollars', he may also not realize that he is the intended inheritor of the will or know whether this inheritor is still living. No such possibility of ignorance occurs when Smith

---

[1] 'The First Person' in S. Guttenplan (ed.), *Mind and Language*, Oxford: Clarendon Press, 1975, pp. 45–65.

says 'I inherit five thousand dollars'. It makes no sense here to speak of
Smith not knowing who inherits the money or whether the inheritor
exists. Hence, Anscombe concludes, the name of a person cannot be
substituted for the pronoun 'I', and the pronoun cannot be regarded as
a referring expression, with the same function as the name. A parallel
argument is given for demonstratives. Smith may say pointing to what
he thinks is behind a screen 'This person is sitting' not knowing who he
is referring to. It may also be possible that there is nothing behind the
screen or that what is there is not a person. But then 'this person'
cannot be substituted for 'I' in 'I am sitting' said by Smith, since neither
ignorance nor non-existence are possible in using the first person form
of the sentence.

There are contexts, however, in which it seems perfectly legitimate
to substitute proper names and demonstratives for the first person
pronoun. By an *address* we shall mean any proper name used to indicate
either the addressee of a token of a sentence, the person for whom it is
intended, or its addressor, the person using it either in speech or writing.
When someone says 'Tom, it is raining outside' the name 'Tom' is being
used as an address to indicate to Tom he is the addressee of 'It is raining
outside' on that occasion. It is clearly not being used to refer the hearer
to some object to be identified of which some predicate is being ascribed,
and hence is not a logical subject. Similarly, when Smith says over the
telephone 'Tom, this is Bill. I am sitting' 'Bill' is an address indicating
for Tom who the speaker is. It no more refers to this speaker than does
'Tom' refer to the hearer. In written communication the same indicating
role is played by the names in the salutation and close of letters and in
the addresses and return addresses on envelopes. For Bill to write
'sincerely, Bill' is not for him to refer to himself in order to ascribe some
predicate; rather, it is for him to indicate who wrote the letter.

When the pronoun 'I' occurs in a sentence preceded by a name it
indicates what is indicated by this address and is a pronoun of laziness.
Contrary to Anscombe's view, in such contexts the name can be sub-
stituted for the pronoun without any alteration of use. Instead of saying
over the telephone 'This is Bill. I am sitting' Bill could say 'This is Bill.
Bill is sitting'. The latter sounds awkward, but this is only due to a
redundancy that the pronoun is designed to avoid. Also, by using the
pronoun we avoid the possibility of the second occurrence of the name
indicating someone other than the speaker. But this in practice would
not occur. If the speaker intended to refer to another, he would use a
form of his name distinct from his own, perhaps supplemented by an
identifying description. Certainly the reasons Anscombe gives for not
substituting a name for the first person pronoun are inapplicable in this
case. Once having named himself with an address there is no possibility
that a speaker not know whom a second occurrence of that same name

indicated and, of course, no possibility that the address failed to indicate anyone.

There are also forms of demonstrative phrases that can be substituted for 'I' while still avoiding Anscombe's difficulties. It is common in formal writing to use such forms of sentences as 'This reviewer thinks that p', 'This writer argues that p', etc., where the phrases 'this reviewer' and 'this writer' are stylistic substitutes for the pronoun 'I'. Like 'I' in the contexts in which names can be substituted, the demonstrative phrase's indicating role is parasitic on an address prefacing or following the discourse, the name of the reviewer, the writer, etc. Though not common, there seems no reason why 'this speaker' cannot also be used as a substitute in spoken sentences. Again, there is no possibility of misidentification: for there to be an utterance there must be a speaker and the demonstrative phrase can only indicate this person.

The fact that substitution is possible does not require our holding that 'I' is synonymous with 'this speaker' or 'the person producing this utterance'.[1] That this is not the case is shown by the fact that whereas 'I am speaking' is a contingent statement 'This speaker is speaking' is analytic. What the substitutability of 'this speaker' for 'I' does seem to show is that the role of the pronoun is to indicate what is indicated by the demonstrative phrase. Its role is that of an address indicating who it is that is producing the utterance.

Both types of substitution denied by Anscombe, that of a proper name and of a demonstrative phrase, thus serve to substantiate her conclusion that the pronoun 'I' is not a referring expression. Where the pronoun is proxy for a name the name has the role of an address, and the demonstrative phrase that can be substituted is also one indicating the addressor. The speaker does not use the pronoun to identify himself, for this presupposes the possibility of misidentification, the speaker's ignorance of the referent, and the non-existence of the referent. Since none of the possibilities essential for reference hold for the use of 'I' Anscombe's conclusion follows.

Southern Illinois University at Carbondale   © D. S. CLARKE, JR. 1978

[1] For a criticism of this view held by Reichenbach see H. N. Castaneda, 'Indicators and Quasi-Indicators', *American Philosophical Quarterly*, Vol. 4 (1967), pp. 85–100.

BLACKBURN ON THE INTERSUBSTITUTABILITY
OF PROPER NAMES

By B. F. KEATING

IN 'The Identity of Propositions' Simon Blackburn defends the view
that co-referential proper names are identical in sense against the
Fregean charge that they are not intersubstitutable *salva veritate* in pro-
positional attitude contexts.[1] The notion of sense employed is Fregean:
two proper names are identical in sense if and only if 'the contributions
they make to the truth-conditions of any whole sentence incorporating
them' are the same (p. 198). But Blackburn's account of the sense of a
name is not Fregean: where N is a proper name, its contribution to the
truth-conditions of a sentence $F(N)$ is simply to "introduce" an object
X such that the truth-condition for $F(N)$ is that $<X>$ satisfies $F(\ )$. I
call such an account 'referentialist'.[2]

The defence proceeds by granting that two names identical in sense
are not intersubstitutable while arguing that we should not expect them
to be. In Blackburn's terms this amounts to an attack on

(1) If two names make the same contribution to truth-conditions, then
they are intersubstitutable *salva veritate* in attitude contexts.

I want to suggest that the considerations Blackburn adduces against (1)
serve better as reasons against the referentialism he endorses.
Blackburn writes:

> To suppose that [sameness in truth-conditions contribution] must
> [license intersubstitutability] would be to legislate that a language cannot
> contain a term which (i) makes such a [referential] contribution to the
> truth-conditions of sentences *and* (ii) is yet applied by speakers to a thing
> *because* they see that thing as having some feature. But this is certainly
> possible (p. 204).

The main premise here is this: (1) implies, for a proper name N, that the
conjunction of

(2) N makes a referential contribution to truth-conditions
and

(3) N is applied to a thing by speakers because they see that thing as having feature F

is impossible. Blackburn is concerned to argue that ((2) and (3)) *is* possible. And he does this by arguing that a name can be applied 'because of some feature of its bearer without that feature playing any role in its meaning' (p. 204).

Now it seems to me that Blackburn is on strong ground when he grants that co-referential proper names are not intersubstitutable *salva veritate* in attitude contexts. But his fellow referentialist who claims otherwise will not be blocked by the above argument. He will agree that, although 'Hesperus' is applied to an item by (some) speakers because they see that item as being the evening star, nevertheless, *something is the evening star* is never part of the truth-conditions contribution of 'Hesperus' to any sentence containing it. The above argument yields no reason to deny the consistency of (1)—(3) and if these are consistent (1) does not imply that ((2) and (3)) is impossible.

Now Blackburn also claims that the intersubstitutability of two proper names in attitude contexts requires 'sameness in the epistemology of their application' (p. 205). So it is natural to read Blackburn as denying (1) on the grounds that sameness in the truth-conditions contribution of two names does not guarantee sameness in the epistemology of their application. This accords with Blackburn's commitment to referentialism. The general idea is that a true report of a propositional attitude 'needs to consider *more* even than the simple semantic identity of components, but also whether the epistemology of their application matches the epistemology of the application of the components in the reported person's expression of his thought' (p. 205, my emphasis); and considering this extra factor involves considering what features of an item the reported person counts as crucial to determining whether a component name applies to that item (see below).

On behalf of the claim that a name can be applied because of some feature of its bearer without that feature figuring in the name's meaning, Blackburn writes:

> A good example of this comes with demonstratives: a conjurer can correctly describe his audience by saying "They don't know that *this* is the same as *this*"—showing some object under different aspects. But a direct semantic account of the demonstrative is obviously right (p. 204).

Why is a direct semantic, or referential, account obviously right? On a 'direct semantic account of the demonstrative', the truth-condition of (what is expressed by) 'This is the same as this' is the same whether or not the object demonstrated is shown under different aspects; the truth-condition is (something like): X is self-identical, where X is the object demonstrated. It seems that the proper reply to Blackburn here is that,

if the direct semantic account is correct, then the conjurer does *not* qualify as expressing what his audience does not know, using 'This is the same as this'. For Blackburn himself would say that what the audience does not know is that the thing with aspect *A* is the same as the thing with aspect *B*. ('. . . I may well be ignorant that Chomolungma is Everest, and this will be because of my ignorance that . . . two sets of features apply to one mountain' (p. 204).) But a truth-condition of *the thing with aspect A is the same as the thing with aspect B* is that something has aspect *A*, which is not a truth-condition of (what is expressed by) 'This is the same as this' on the direct semantic account of the demonstrative.

Blackburn's position is that *only some* sentences of the form

(4)   ———is the same as——— \

(where the blanks are to be filled by names referring to what the conjurer demonstrates) express beliefs of the audience. And he says that 'to identify the sense of a particular sentence is . . . to recognize which set of truth-conditions it creates out of the senses of the component terms' (p. 198). The sense of a sentence is also regarded as 'the determinant of a set of truth-conditions' (p. 198). Now, among other things, beliefs are true or false and hence possess truth-conditions. But then identifying the belief expressed by a sentence is identifying its sense, i.e., identifying 'the determinant of a set of truth-conditions' associated with that sentence. The direct semantic account makes it especially difficult to distinguish the sentences of this form which express beliefs of the audience from those which do not. For, assuming that the direct semantic account is correct for names, all sentences of this form have the *same* truth-conditions.

This suggests to me a reason why the direct semantic account is false. But Blackburn claims rather that 'there is no immediate reason why the thing which is believed . . . should be identical with the thing introduced by the theory of truth' (p. 182)! Hence his attack on (1). Yet Blackburn also writes: '. . . the . . . problem, of individuating objects of the understanding (or, in more modern dress, giving truth-conditions for sentences using propositional attitude constructions) is apparently as dark as ever' (pp. 204–5). This is not surprising if the determinant of a set of truth-conditions is not identical with that toward which one has a propositional attitude in the first place! It seems to be only Blackburn's endorsement of referentialism, and his consequent commitment to the view that all the sentences of our form are identical in truth-conditions, which leads to his separation of 'the determinant of a set of truth-conditions' from 'the object of the understanding', and hence to his problem about giving identity conditions for the latter.

Suppose the audience gives the name 'Lesley' to the object demonstrated at the first occurrence of 'this' because they see it as female, and

gives the name 'Chris' to the object demonstrated at the second occur-
rence of 'this' because they see it as male. Blackburn makes the following
general claim about determining the sense of a term: 'To determine the
native sense of a term [is] precisely to map which features of the world
the speakers take as crucial to determining the truth or falsity of sentences
involving it' (p. 198). However, Blackburn must not intend that we
count proper names as terms. For, if proper names are terms, then,
assuming that a sentence 'creates' a set of truth-conditions out of
the senses of its component terms, the thesis that all the sentences of the
aforementioned form (4) are identical in truth-conditions implies that all
the component names are identical in sense. And, if Blackburn's general
claim about sense is true, this implies that the audience does not regard
settling whether something is both male and female as crucial to deter-
mining the truth or falsity of, e.g., 'Lesley is the same as Chris'. For
they do not regard settling this as crucial to determining the truth or
falsity of 'Lesley is the same as Lesley' which is supposed to have
the same truth-condition as 'Lesley is the same as Chris'. But it is
simply not true that the audience will regard the same features of the
world as crucial to determining the truth of falsity of *all* sentences of the
form (4).

The statement about determining the native sense of a term quoted
in the previous paragraph does not rest well with the referentialist view
of proper names. Blackburn speaks of a man's applying 'Everest' to a
mountain because it has particular features seen from the south. Among
these features, we may suppose, is *being snow-capped*. And perhaps such
a man will not regard whether or not some snow-capped thing is tall as
'crucial to determining the truth or falsity' of 'Everest is tall'. But another
of these features, surely, is *being a mountain*. To defend the referentialist
view, however, Blackburn must deny that our man regards whether or
not some mountain is tall as crucial to determining the truth or falsity of
'Everest is tall'.

Clearly, then, Blackburn does not intend the above statement about
determining the native sense of a term to apply to proper names. Let us
suppose that the statement does not so apply. Blackburn claims that the
'because' in 'because they see that thing as having feature F' is 'entirely
causal'. I cannot explicate this. But if it implies that 'the truth of what is
asserted is . . . not a matter of how it is with a thing satisfying a descrip-
tion or cluster of descriptions, even if they served to introduce the name
in the first place' (p. 203), then Blackburn has simply no account at all to
offer of the truth-conditions of the sentences of the form (4). *X is self-
identical* works in such an account only if we understand X to be (say) the
object demonstrated by the conjurer, which leaves us with the question
whether or not it is a truth-condition of sentences of our form that there
is a conjurer. On the direct semantic, or referential, account—which

REFERENCE, TRUTH-FUNCTIONALITY AND CAUSAL SENTENCES

By A. J. DALE

I

MY aim in this paper is to examine the relationships that exist between reference, truth-functionality and causal sentences. It gives support to Quine's thesis that contexts which are referentially transparent and admit the substitution of logical equivalents *salva veritate* are truth-functional. An argument of Quine's for this conclusion has recently been attacked.[1] Because of the use to which Quine's argument and conclusion have been put by both Anscombe[2] and Davidson[3] in their discussion of causality it is of some importance to see how much of Quine's argument and conclusion can be rescued from such criticism[4].

Quine's original argument relied on the purported logical equivalence between a statement 'p' and the statement '$\hat{x}\{x=\Lambda.p\}=\{\Lambda\}$'. The counter-arguments referred to above all deny this logical equivalence on the grounds that the latter statement entails the existence of at least one class whereas the former does not. In order to throw the main issues into relief I shall use a different although related argument which does not rely on classes. Suppose '$\Phi(\quad)$' is a context in which a sentence may be embedded and which also satisfies the following conditions:—
'$\Phi(\quad)$' allows the substitution of logical equivalents *salva veritate* and is also referentially transparent. (From now on I shall refer to such a context as an *extensional* context). Then the following will all have the same truth-value if 'p' and 'q' have the same truth-value and do not include semantic terminology:—

    i. $\Phi(p)$
    ii. $\Phi(\text{'}p\text{' is true})$
    iii. $\Phi(\text{the truth-value of '}p\text{' is truth})$
    iv. $\Phi(\text{the truth-value of '}q\text{' is truth})$
    v. $\Phi(\text{'}q\text{' is true})$
    vi. $\Phi(q)$.

For i, ii, and iii are logically equivalent (as are iv, v and vi) and iv results from iii by substituting co-referring expressions. Thus, the argument runs, '$\Phi(\quad)$' is a truth-functional context.

---

[1] Quine's argument can be found in 'Three Grades of Modal Involvement' *The Ways of Paradox*, p. 161. The arguments against it are found in Cummins and Gottlieb 'On an Argument for truth-functionality', *American Philosophical Quarterly*, 1972, pp. 265–269 and Lycan, 'The Extensionality of Cause, Space and Time', *Mind*, 1974, pp. 498–511.

[2] Anscombe, 'Causality and Extensionality' *Journal of Philosophy*, 1969, pp. 152–159.

[3] Davidson, 'Causal Relations', *Journal of Philosophy*, 1967, pp. 691–703.

[4] Since this article was presented for publication in ANALYSIS arguments similar to that contained in Section II have been published by C. McGinn, *Mind* 1976, and B. Taylor in *Truth and Meaning*, edd. Evans and McDowell, 1976. Neither, however, made use of the same truth-conditions for the description operator as those given in my article.

This rather simple argument will serve to illustrate the analogous but more complicated arguments of Quine, Davidson and Anscombe. It shows that we do not have to make reference to the null class (Quine), the universal class (Davidson), a logically non-empty class (Anscombe) or numbers (Anscombe). Since it is an analogous argument it is open to analogous criticisms. They would run as follows. Either '*the truth-value of "p" is truth*' entails the existence of a truth-value, in which case it is not logically equivalent to '*p*' which does not entail the existence of a truth-value. Or '*the truth-value of "p" is truth*' does not entail the existence of a truth-value (e.g. by a contextual definition of '*the truth-value of "p" is truth*' as ' "*p*" *is true*') in which case '*the truth-value of "p"* ' does not refer and so is not admissible as a candidate for substitution *salva veritate*. Whilst I believe that it is possible to counter these criticisms directly, it would require a radical scrutiny of certain key notions in the philosophy of logic—contextual definition, reference to abstract objects, the function of identity statements, ontological commitment, etc. Instead of tackling this daunting task I shall present a number of arguments which, if taken together, would establish a great deal of Quine's thesis. Certainly enough is established to rescue Davidson's and Anscombe's positions.

Since the arguments are independent of each other, an objection to one of them will still allow the conclusions established by the others. Section II will consider substitution of true sentences, section III of false ones.

<div align="center">II</div>

Suppose '$F(a)$' and '$G(a)$' are true sentences in which 'a' occurs purely referentially in Quine's sense, i.e. are such that they allow co-referring expressions to be substituted *salva veritate*. The argument I give below establishes that if '$\Phi(\ )$' is an extensional context then '$\Phi(\ )$' will allow the substitution of '$F(a)$' for '$G(a)$' *salva veritate*. The argument depends crucially on the logical equivalence between '$F(a)$' and '*a is the object x such that x is identical to a and $F(x)$*'.[1] I shall abbreviate '*the object x such that . . . x . . .*' to '$\imath x(\ldots x \ldots)$'. It is not intended that the introduction of the latter symbol should carry with it any doctrine of definite descriptions; I use it simply because it is familiar and some abbreviation is necessary to aid lucidity. To support this logical equivalence it is clear that if '$F(a)$' is true so is '$a = \imath x(x = a . F(x))$' and also that if '$a = \imath x(x = a . F(x))$' is true then so is '$F(a)$'. What might be

---

[1] Follesdal has a related argument to show that causal distinctions disappear in a certain formal system. Unfortunately, it is open to the same objections as Quine's original argument since it relies on the logical equivalence of p and $\imath x \{x = y . p\} = y$. Here he relies on the theory of definite descriptions, treating y as a variable, so blocking a substitution later in the proof. (D. Follesdal, 'Quantification into causal contexts', *Reference and Modility*, ed. L. Linsky 1971, p. 55.)

thought to provide some difficulties would be the case where '$F(a)$' is false and thus '$a=\imath x(x=a.F(x))$' of doubtful truth-value. There are at least two positions that may be taken up if this is thought to give rise to objections to the logical equivalence. Firstly, and this is my own position, that an identity statement containing an expression that does refer and one that does not is false: it is simply false that the king of France is the editor of ANALYSIS. This position does not commit me to the view that *all* subject-predicate statements have a truth-value even when the subject term fails to refer but only that some do. For those who still think of such identity statements as truth-valueless there are Linsky's[1] arguments which, applied to this case, would show that the relation between the two statements is one of logical equivalence even though one statement could be false and the other truth-valueless.

Now if '$F(a)$' and '$G(a)$' are both true '$\imath x(x=a.F(x))$' and '$\imath x(x=a.G(x))$' both refer to the same object, namely what 'a' refers to. Using this fact and the above logical equivalence it can be seen that the following all have the same truth-value:—

   i. $\Phi(F(a))$
   ii. $\Phi(a=\imath x(x=a.F(x)))$
   iii. $\Phi(a=\imath x(x=a.G(x)))$
   iv. $\Phi(G(a))$

This argument is not of course sufficient to establish Quine's thesis but it is sufficient for Davidson's and Anscombe's purposes. Consider the example of Davidson's '*Smith's death was caused by the fall from the ladder*' and the related sentence '*The fact that Smith fell from the ladder caused it to be the case that Smith died*'. Davidson then uses the fact that such sentences allow the substitution of co-referring expressions and logical equivalents *salva veritate* to show that the latter sentence cannot represent the logical form of the former sentence since the latter sentence would then have to be truth-functional which it is not. Anscombe, on the other hand, argues that all this shows is that statements of causality do not allow the substitution of co-referring expressions *salva veritate*. Both Davidson and Anscombe make use of Quine's argument as well as his thesis in order to support their claims and so they are both vulnerable to the criticisms made of Quine's argument. Both their arguments can be reinstated, however, by using instead of Quine's argument the argument of this section. For suppose that Smith, as well as falling off the ladder, was also a philosopher. Then the argument shows that '*Smith was a philosopher*' should be substitutable *salva veritate* for '*Smith fell off the ladder*' in the above causal context, if the causal context is indeed extensional. Since it is not so substitutable both the Davidson and Anscombe arguments go through.

[1] Linsky, *Referring*.

III

By an argument analogous to that contained in II, it can be established that if '$\Phi(\ )$' is an extensional context then any false sentence containing a purely referring expression can replace any false sentence containing a purely referring expression in that context *salva veritate*. This is a more general result than that of Section II since here the referring expressions need not refer to the same object. Suppose that '$F(a)$' and '$G(b)$' are both false and 'a' and 'b' occur purely referentially. Then the following all have the same truth-value:—

i. $\Phi(F(a))$
ii. $\Phi(a=\imath x(x=a.F(x)))$
iii. $\Phi(\imath x(x=a.\sim F(x))=\imath x(x=a.F(x)))$
iv. $\Phi(F(a).\sim F(a))$
v. $\Phi(G(b).\sim G(b))$
vi. $\Phi(\imath x(x=b.\sim G(x))=\imath x(x=b.G(x)))$
vii. $\Phi(b=\imath x(x=b.G(x)))$
viii. $\Phi(G(b))$.

Each line of the above is the result of substituting logical equivalents (i—ii, iii—iv, iv—v, v—vi and vii—viii) or co-referring expressions (ii—iii and vi—vii). That the statements in iii and iv (v and vi) are logically equivalent can be shown briefly. Suppose that $\imath x(x=a.\sim F(x))=\imath x(x=a.F(x))$ and $F(a)$. Then $a=\imath x(x=a.F(x))$ and so $\imath x(x=a.\sim F(x))=a$, thus $\sim F(a)$. The last statement contradicts the supposition and so $\sim F(a)$. *Mutatis mutandis* we can show all the other entailments necessary to provide the logical equivalence.

The argument does rely on the logical equivalence between '$F(a).\sim F(a)$' and '$G(b).\sim G(b)$'. There are some, perhaps, who would deny that these two contradictions are logically equivalent and clearly for them the above argument will not work.[1] But this is not the place to discuss this issue and I have dealt with this question elsewhere.[2]

Combining the arguments of this section with the argument of the last I have shown that if '$\Phi(\ )$' is an extensional context then any sentence containing 'a' purely referentially can be substituted for any other sentence of the same truth-value containing 'a' purely referentially in '$\Phi(\ )$' *salva veritate*.

IV

Before the main argument of this section I should like to discuss an example which shows up what I believe to be the central weakness of Davidson's argument. Consider the following pair of sentences:—

---

[1] There are indications of this position in Geach 'Entailment', *Proceedings of the Aristotelian Society*, 1958, pp. 157–172 and later papers, although his argument there is not specifically against a contradiction entailing another *contradiction*.

[2] Dale, 'Geach on Entailment', *Philosophical Review*, 1973, pp. 215–219.

'Smith cannot join the police force because his height is five feet' and *'Smith's height is the cause of his rejection by the police force'*. It seems clear that in these cases we cannot substitute for *'Smith's (his) height'* a co-referring expression, *'five feet'* or *'one and two-thirds yards'* for example, *salva veritate*. It would seem *prima facie* that Davidson's claim that co-referring terms can replace each other in a causal sentence *salva veritate* is simply false and thus Anscombe's claim that this is indeed the case is justified. It could be objected that neither an abstract object (Smith's height) nor a relationship between abstract objects (the identity between Smith's height and five feet) can be the cause of anything. Such an argument, however, would not save Davidson for his argument depends precisely on the substitution of the identity '$\hat{x}\{x=\Lambda.p\}=\{\Lambda\}$' which is an identity between abstract objects for 'p' in a causal statement.

Indeed, it would seem that even if the logical equivalence between 'p' and the class identity statement is granted Davidson must show why he thinks that causal sentences are such that they allow the substitution of such logical equivalents *salva veritate*. Hopefully, my argument of section II could replace Davidson's argument *in toto* and free it from this objection also.

Leaving aside the question of whether a causal statement can contain a sentence expressing an identity between abstract objects, there are certainly many contexts that can. Let us consider identity statements of the form '$a=b$' where both 'a' and 'b' are expressions denoting finite numbers. (Such an identity will henceforward be called a *numerical identity*.) Then one can show that if '$\Phi(\quad)$' is an extensional context then '$\Phi(\quad)$' will allow the substitution of one numerical identity for another of the same truth-value *salva veritate*. Although this conclusion may seem fairly limited, it should be remembered that for those who like to unify their discourse there are such diverse ways of referring to numbers as *'Smith's height in feet'* and all the other results of transforming statements of "impure" number to statements of "pure" number.

Let us assume, then, that '$a=b$' and '$c=d$' are two true numerical identities. Then, since a number k can always be found such that $a=c+k$ and $b=d+k$, the following three propositions all have the same truth-value:—

  i. $\Phi(a=b)$
  ii. $\Phi(c+k=d+k)$

(since '$c+k$' and 'a' are co-referring terms and similarly 'b' and '$d+k$')

  iii. $\Phi(c=d)$

since '$c=d$' is logically equivalent to (entails and is entailed by) '$c+k=d+k$'.[1]

---

[1] If it is objected that we may not be able to specify k, I may add that there always is a specification, namely '$a-c$'|

Secondly, let us assume that '$a=b$' and '$c=d$' are two false numerical identities. Then, there are numbers m and n such that $m\neq o$, $a=mc+n$ and $b=md+n$. Then the following three propositions all have the same truth-value:—

    i. $\Phi(a=b)$
    ii. $\Phi(mc+n=md+n)$

(since 'a' and '$mc+n$' are co-referring individual descriptions and similarly 'b' and '$md+n$')

    iii. $\Phi(c=d)$

since, when $m\neq o$, '$mc+n=md+n$' is logically equivalent to '$c=d$'.

Thus we have a limited truth-functionality theorem:—a context that allows the substitution of referring terms and logical equivalents *salva veritate* will allow the substitution of one numerical identity for another of the same truth-value *salva veritate*.

V

The success of the limited conclusion of the last section depends upon there being a two place predicate expression which could be applied truly to the pairs a, c and b, d and which provided a description of a in terms of c (b in terms of d) that could, as it were, be turned around to give a description of c in terms of a (d in terms of b).

If we tried to extend the proof of the last section to any identities whatsoever, we should need to be sure that such predicate expressions always existed linking pairs of objects in the desired way. But how can we be sure that such expressions always exist?

Certainly, in some cases such expressions exist. Take, for example, the two true identities '*Paris is the largest city on the Seine*' and '*France is the largest country in Western Europe*'. The expression '——*is the sole capital of the country*——*and*——*only*' where the last two blanks are filled by the same name will perform as such a suitable expression for these two identities, since all the following will have the same truth-value if '$\Phi(\quad)$' is an extensional context:-

    i. Φ(*Paris is the largest city on the Seine*)
    ii. Φ(*the sole capital of the country France and France only is the largest city on the Seine*)
    iii. Φ(*France is the sole country whose sole capital is the largest city on the Seine*)
    iv. Φ(*France is the largest country in Europe*).

The transition from i to ii and from iii to iv is obtained by the substitution of co-referring expressions and the transition from ii to iii by the substitution of logical equivalents.

No doubt with sufficient ingenuity we could find expressions which

would work similarly for other identities but there is another argument which applies quite generally to all pairs of true identities. Suppose that '$a=b$' and '$c=d$' are two true identities and that '$\Phi(\quad)$' is an extensional context. Then the following have the same truth-value:—

    i. $\Phi(a=b)$
    ii. $\Phi(a=b.c=c)$
    iii. $\Phi(a=a.c=d)$
    iv. $\Phi(c=d)$

In the above iii results from ii by the substitution of co-referring expressions. ii results from i by the substitution of the logically equivalent '$(a=b.c=c)$' for '$a=b$' and iv from iii by a similar substitution. If it is objected that '$c=c$' is not a necessary truth since it entails the existence of c there is a more tedious proof available using '$(c=c) \vee (\sim(Ex)x=c)$' and '$(a=a) \vee (\sim(Ex)x=a)$' as necessary truths.

<h1 style="text-align:center">VI</h1>

A slightly more complicated argument has to be used for any pair of false identities. Suppose 'a', 'b', 'c', 'd' are referring terms that have referents and that '$a=b$' and '$c=d$' are false identities. Then, if '$\Phi(\quad)$' is an extensional context all of the following have the same truth-value:—

    i. $\Phi(a=b)$
    ii. $\Phi(c=c \supset a=b)$
    iii. $\Phi(\imath x(x=c.x\neq d)=c \supset a=b)$
    iv. $\Phi(c\neq d \supset a=b)$
    v. $\Phi(c\neq d \supset \imath x(x=a.x=b)=b)$
    vi. $\Phi(c\neq d \supset \imath x(x=a.x=b)=\imath x(x\neq a.x=b))$
    vii. $\Phi(c\neq d \supset (a=b.a\neq b))$
    viii. $\Phi(c=d)$

iii is obtained from ii and vi from v by substituting co-referring expressions. The other members of the sequence have relied on the substitution of logical equivalents which themselves depend on '$c=c$' being a necessary truth (ii from i), '$(a=b).(a\neq b)$' being a contradiction (viii from vii) and 'Fa' being logically equivalent to '$\imath x(x=a.Fx)=a$' (iv from iii, v from iv, vii from vi). As in the argument in Section VI it may be objected that '$c=c$' is not a necessary truth since it entails the existence of c. Again a similar but much more complicated argument can be given using '$c=c \vee \sim(Ex)x=c$' and related propositions as necessary truths.

<h1 style="text-align:center">VII</h1>

All that remains now is to combine the different arguments given above to establish that a context which is extensional will allow the substitution of any sentence containing a purely referring expression

for any other sentence containing a purely referring expression *salva veritate*. For the following have the same truth-value if '$\Phi(\ )$' is extensional and '$F(a)$' and '$G(b)$' have the same truth-value:-

  i. $\Phi(F(a))$
  ii. $\Phi(a=\imath x(x=a.F(x)))$
  iii. $\Phi(b=\imath x(x=b.G(x)))$
  iv. $\Phi(G(b))$

for i and ii (iii and iv) are logical equivalents (Section II) and ii and iii have been shown to have the same truth-value by the arguments of IV—VI.

 Although I have not demonstrated the full Quinean thesis, for certain types of sentences—'*unicorns do not exist*', '*all ravens are black*'—have not been mentioned, enough has been shown I believe to support Quine's contention that the blame for non-truth-functionality rests on the manner in which we refer to things. Certainly, sufficient has been produced to repair the arguments of Davidson and Anscombe. The suspicion that a trick (Anscombe's terminology!) has been played on us by Quine's original neat argument appealing to the identity of classes should now be allayed by the variety and number of the different arguments displayed in this paper although it might, I suppose, be thought that all I have done is to replace one trick by a whole box of tricks.

The University of Hull   © A. J. DALE 1978

DESCARTES, FRANKFURT AND MADMEN

By Steven DeHaven

IN the *First Meditation* Descartes introduces a brief discussion of mad-
men. In *Demons, Dreamers and Madmen*[1] Harry Frankfurt describes
this as follows:

> The fact that error may arise out of madness suggests that he should be
> suspicious of all of his opinions until he can establish that he is not insane.
> He does not, however, attempt to establish his sanity, or even provide a
> procedure for doing so.
>
> Instead he rather abruptly dismisses the suggestion. He asserts, indeed,
> that he would be mad to compare himself with the lunatics he has de-
> scribed (p. 37).

It will be my contention that the suggestion is not, philosophically
speaking, dismissed. In Descartes' view it is of the same epistemological
force as the more famous suggestion that one might be dreaming. One
must keep in mind that Descartes faces a twofold task. On the one hand
he must justify doubt, but on the other he must overcome "psycho-
logical" features of his readers which might prevent them from accept-
ing good reasons for doubt. In this case the move from madmen to
dreamers is made, I think, primarily because Descartes believes it will
be more persuasive.

As Frankfurt notes the *Meditations* have a dialectical character (p. 79).
This is to be expected since in the course of meditation one often pro-
duces pro and con arguments. Though it is slightly artificial I shall view
the *First Meditation* as a sort of dialogue between two individuals—the
meditator and the plain man. (An analogue of the meditator is Eudoxus
in 'The Search After Truth'.) It is the task of the meditator to carry
through the overall project of the *Meditations*. Since the meditator is the
principal spokesman for Descartes himself, it is of some importance to
decide to which of the two characters to attribute a given remark.

The question of how the existence of one's hands and one's body
could be doubted is raised. Then the discussion proceeds as follows (I
have inserted the italics):

> Unless indeed I likened myself to some lunatics, whose brains are so upset
> by persistent melancholy vapours that they firmly assert they are kings,
> when really they are miserably poor; or that they have a head of pottery,
> or are pumpkins, or are made of glass; *but then they are madmen, and I*
> *should appear no less mad if I took them as a precedent for my own case.*
>
> A fine argument! As though I were not a man who habitually sleeps
> at night and has the same impressions (or even wilder ones) in sleep as
> these men do when awake!

[1] Page references to Frankfurt are to this book. The English translations of Descartes
are those of Anscombe and Geach.

The italicized lines are, I suggest, to be attributed to the plain man, the others to the meditator. In speaking of an abrupt dismissal Frankfurt is in effect attributing the italicized line to the meditator. But if my structuring is correct then the attribution of the dismissal to Descartes is at best dubious. For what Descartes is doing is recognizing that the suggestion is one which the plain man will find unpersuasive. He recognizes that there are individuals who would say: 'Descartes, oh yes, he's that silly chap who asked us to think that we are mad'.

The primary support for my structuring comes from the way in which the second cited paragraph opens. The remark 'praeclare sane', translated by Anscombe and Geach as 'A fine argument' and by Frankfurt as 'How eminently reasonable', seems to have an ironic, indeed sarcastic, tone. That this is so is indicated by the punning use of 'sane', where insanity was the previous topic. Frankfurt's own translation seems to concede at least this point. If this is so, then one must answer the question what the scope of the remark is. Does it, for example, apply to the whole of the preceding paragraph? Or does it apply to the portion I have cited? Neither is, I think, the case. For Descartes continues by way of comparing the state of the dreamer to that of the madman. The use of 'as though I were not . . .' suggests that what is being dismissed by the meditator, hence by Descartes, is the abrupt dismissal I have attributed to the plain man. The plain man has tried to evade an argument that all "physical object statements" are doubtful, but he fails.

That the suggestions about madness and about dreaming are not meant to differ epistemologically is indicated by the way in which both are characterized. To dream is to be mad on certain occasions. Both are characterized as involving perceptually bizarre experiences and judgments. Frankfurt recognizes this (p. 40), but nonetheless persists in holding to the abrupt dismissal view. As I have suggested this leaves him no way to account for the force of 'praeclare sane'. And again, insofar as dreaming and madness are states of the very same sort, one is, in contemplating that one might be dreaming, contemplating in effect that one might be mad.

In *Descartes: A Study of his Philosophy* Kenny suggests that the question of how one knows one is not mad is not pursued; possibly, he says, so as to avoid giving offence to the reader (p. 29). However the offence is important only in that it is related to unpersuasiveness. And given the sameness of madness and dreaming, Descartes is pursuing the question by means of pursuing the question about dreaming. At this point just as there are no certain indications that one is not dreaming there are no certain indications that one is not mad. In the end, that is, after the completion of the project of the *Meditations* Descartes would give just the sort of unexciting answer to a question as to how one knows one is not mad as to how it is that one knows one is not dreaming. I doubt if

Descartes thought seriously about the sort of "philosophic" questions (which both Frankfurt and Kenny have in mind) as to whether a madman can or cannot know he is mad, as to whether a sane person can know he is sane. Such "analytic" questions were typically viewed by Descartes as not worth serious philosophic attention.

Of course there is a sort of problem about what Descartes has in mind in speaking of madness. Frankfurt leans toward an account in which madness is a loss of a capacity to distinguish between reasonable and unreasonable judgments; it involves a loss of rationality (p. 38). However the texts I have cited, with their emphasis on perception, do not support so strong an interpretation. And indeed it is doubtful if Descartes can consistently accept such an interpretation. Since men are an image of God, understanding, albeit finite, is part of their essence. It would seem then that no man could lose his rationality in the sense contemplated by Frankfurt.

Are there any other factors which enter into the choice of arguments? Each reader of the *Meditations* is supposed to draw upon his own resources rather than upon knowledge of others or knowledge which they purportedly have. From this standpoint there is an asymmetry between the dream argument and the madman argument. In the former case a sceptical conclusion is justified via the use of a first-person factual report. I have had dreams. But in the madman case one begins with the claim that some men are mad. This does not seem, or at least it might not to the plain man, to lend support to the claim that I might be mad. I have suggested that Descartes allows that it does. But as we have seen, his route to the possibility of madness is, for reasons of persuasiveness, via the similar and familiar state of dreaming.

University of Alberta   © STEVEN DEHAVEN 1978

THE IRRELEVANCE OF THE FREE WILL DEFENCE

By Steven E. Boër

IT is traditional, when discussing the theological Problem of Evil, to distinguish between "natural" and "moral" evils and to employ somewhat different strategies in dealing with these two species. Natural evil has proved difficult to rationalize, whereas the problem posed by the existence of moral evil has seemed to many to succumb to some version of the classical Free Will Defence. Indeed, the Free Will Defence has appeared so promising that Plantinga[1] has even suggested that it extends to natural evil as well, the idea being that so-called natural evils might be regarded as moral evils ascribed to Satan and his crew of fallen angels. The purpose of this note is to point out that the Free Will Defence is simply irrelevant to the Problem of Evil in its most poignant form, and that it is the problem of natural evil which threatens to swallow that of moral evil rather than the other way around.

The Free Will Defender is concerned to argue that the following two propositions are logically incompatible:

(1) God creates a world W in which persons have free will.

(2) Persons in W do no evil.

Since the inconsistency is a logical one, God's omnipotence is not compromised by the fact that He cannot create a world of free persons who do no evil. And since obligation presupposes ability, God's perfect goodness is supposedly reconciled with the existence of moral evil.

The problem with the Free Will Defence can be traced to a fatal ambiguity in (2), which can be read either as (2a) or as (2b):

(2a) Persons in W are morally perfect agents.

(2b) Persons in W perform no actions having morally objectionable consequences.

In other words, (2) blurs the crucial distinction between the moral evaluation of *agents* and the moral evaluation of *actions*, the former being a matter of motives and intentions and the latter being a matter of consequences. All that the Free Will Defence really shows is that (1) is inconsistent with (2a), i.e., that God (logically) cannot create a world in which persons are free but never make morally reprehensible choices or form morally objectionable intentions. Nothing whatever follows anent (2b). Yet (2b) is precisely the proposition whose apparent falsity underlies the ordinary man's qualms about God's goodness. What bothers us,

---

[1] A. Plantinga, "The Free Will Defence", in Max Black (ed.) *Philosophy in America* (Ithaca, N.Y.: Cornell University Press, 1965).

e.g., about the existence of a man like Hitler is not that he had evil motives, made evil choices, or formed evil intentions, but that God stood by while Hitler *successfully carried out* many of his appalling designs!

To see that (2b) is untouched, we need only reflect on the fact that free will has nothing intrinsically to do with the question of whether or not a given agent *succeeds* in doing what he "willed" to do, unless the intended action pertains to certain kinds of immediate control over his own body. Suppose *A* decides to murder *B*. *A* picks up a revolver, loads it, points it at *B*'s head and squeezes the trigger. If *A* were unable to coordinate his bodily motions so as to grasp the gun, load it, point it, etc., then his free will might be compromised in the sense that he might be prevented from *trying* to shoot *B*. But suppose the gun's mechanism is rusty and the trigger refuses to budge in spite of *A*'s vigorous squeezing. Here there is manifestly no impediment to *A*'s free will: in grasping the gun, aiming it and squeezing the trigger *A* has done something which counts as *trying to shoot B*. *A* no longer qualifies (if indeed he ever did) as a morally perfect agent, but so far he has performed no action leading to morally bad consequences. In short, freedom of the will is freedom of opportunity: it is a licence to choose and try, not a warranty of success.

Thus we are immediately thrown back upon the problem of *natural* evil. God may be guiltless in permitting his creatures to *try* bringing about evil consequences, for preventing this might violate their free will. But God cannot be guiltless in having so designed nature that these evil consequences often *follow causally* from the events (whatever they may be) which constitute trying. For this extension of the causal chain could be prevented without in any way compromising anyone's free will. It should be stressed that this objection is completely independent of any particular theory about the mechanics of free choice: it applies equally whether one is a hard determinist, soft determinist, contracausalist, or whatever. For the objection centres not on choosing and trying but on the natural consequences of choosing and trying, which lie outside the purview of "free agency".

A number of half-hearted rejoinders come to mind at this point, only two of which merit any mention. One might reply that tampering with the relevant causal chains would force God to create a horrendously complex and chaotic universe, and that such a universe would be "bad" in some sense which outweighs the alleged decrease in pain and suffering on the part of the innocent. The answer is simple: if, as is traditionally assumed, God's omniscience includes foreknowledge of all things, then He can (in virtue of His omnipotence) arrange for appropriate coincidence miracles to obviate the evil consequences which would otherwise ensue from attempted wrong-doing. And, by definition, coincidence miracles involve no change in the Laws of Nature themselves. The resulting universe, far from being chaotic, would be

NOTES

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr. Christopher Kirwan, Exeter College, Oxford OX1 3DP. Contributors are asked to note the following requirements.

Articles should normally not exceed 3,000 words in length. Occasionally, however, longer contributions can be accepted;

. They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope and international reply coupons of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

APOLOGY

The Editor apologizes for the title of a recent article, 'Lies, damned lies, and Miss Anscombe', which he now acknowledges was liable to be construed as offensive.

REPORT ON *ANALYSIS* "PROBLEM" NO. 16

By RONALD J. BUTLER

> If Brown in an ordinary game of dice hopes to throw a six and does so, we do not say that he threw the six intentionally. On the other hand if Brown puts one live cartridge into a six-chambered revolver, spins the chamber as he aims it at Smith and pulls the trigger hoping to kill Smith, we would say if he succeeded that he had killed Smith intentionally. How can this be so, since in both cases the probability of the desired result is the same?

There were twenty-three entries, from five countries. The overall standard was commendably high: it is regrettable that space does not permit publication of more entries.

Most contributors assumed that the cases presented in the problem are *as* asymmetrical as described: this mistake was positively encouraged in setting up the problem! Having made this assumption, far too many entrants argued that aiming a six-chambered gun at somebody raises questions of moral responsibility, and that this accounts for the disparity. Nonsense! As Lawrence Carleton of the University of Minnesota neatly remarked, 'Morality is irrelevant, as is the law: Brown could just as intentionally have shot a beer bottle off a wall after setting the revolver spinning.' Surprisingly few who were led in this direction queried the assumption that an *ordinary* game of dice raises no questions of responsibility: that depends upon where and under what circumstances one plays the game. Gambling saloon proprietors *claim* to conduct ordinary games of dice.

Several entrants confidently argued that I had merely set a probability problem, and most of these assumed that if Brown were using a single-chambered gun he would shoot with 100% accuracy! But as Jacek Hołówka remarked, 'An inefficient murderer is still a murderer.' Again Lawrence Carleton laid this argument to rest with the words, 'Probability is irrelevant: Brown still would have killed Smith intentionally'

had the revolver been eight-chambered . . .'. Some thought that probability enters in an eccentric manner: if one intends to do something one attempts to maximise the probability of that event's occurring. Not so: procrastinators may be full of intentions, whether good, bad or indifferent.

Two highly commendable entries had to be disqualified. Alan Morrison of Stourbridge, Worcester (who has taken no courses in philosophy) and Jacek Hołówka of Warsaw were both knocked out by the count.

The prize is awarded to David Ross of Swarthmore College. The expiring Smith makes one telling point after another against Brown: for Ross to have couched this in an elegant dialogue well within the word-limit is no mean feat. The runner-up is Eric Kraemer of the University of Nebraska at Lincoln, who, alive and well, otherwise argues in the manner of Ross's Smith. I have asked that the contribution E. J. Lowe of Ripon, Yorkshire, also be printed, because despite the fact that he commits one of the mistakes mentioned above I have the greatest admiration for his attempt to develop a logic of intentional statements.

HE LOADS THE GUN, NOT THE DICE

By DAVID ROSS

BROWN invites Smith to his flat for a cup of tea and an ordinary game of dice. Brown hopes to throw a six, does so, then claims to have done it intentionally. Smith disagrees. Brown, ever testy, becomes upset. He places one live cartridge into a six-chambered revolver, spins the cylinder as he aims at Smith, and pulls the trigger. The gun fires, mortally injuring Smith.

'There,' says Brown to the dying Smith. 'I just shot you intentionally. By parallel reasoning, I threw the six intentionally.'

'Bad analogy,' gasps Smith.

'What's the difference? My probability of success in each case was the same, and I certainly intended to roll the six just as much as I intended to shoot you.'

' "The most explicit expression of intention is by itself insufficient evidence of intention" ', quotes Smith. 'And as far as probabilities are concerned—well, to begin with, the odds were *not* the same; how good a shot are you? Don't bother to answer, it's unimportant. Probability, or at least its quantitative aspect, has nothing to do with intention. Do we ever assign numbers to degrees of intent?'

'Some acts are considered more intentional than others.'

'Sure, but they're not compared numerically; yet, if intention were a function of probability, it would be possible to set up a *ceteris paribus* situation in which one act was, say, twice as intentional as another. Probability plays an eccentric role in intention: the earmark of an intentional act is the perpetrator's ability to *alter* that probability. Thus, the mark of intentionality in the shooting was the fact of your loading, aiming, and firing the revolver—all actions under your control.'

'But', asks Brown, 'doesn't the fact that I didn't have to throw the die at all evidence some control on my part?'

'Your decision to throw the die only evidenced your willingness to enter into the dice game', answers Smith. 'It wasn't a move *in* that game, and had no influence over the game's outcome.'

'Sure it did. If I hadn't thrown the die, it wouldn't have shown a six.'

'When you win a chess game, do you attribute your victory to the decision to play? Obviously not; you attribute it to the moves you made.'

'So an intentional act must occur—in a sense—as the outcome of a game where there is some skill involved beyond the decision to play. Then answer me this: couldn't the loading, aiming, and firing of the revolver—as with the decision to roll the die—be viewed as the entrance into a sort of "shooting game"?'

'Certainly, with the actual shot—regardless of the ultimate destination of the bullet—corresponding to the outcome of a six on the die. But the gunshot is clearly no more intentional than the fact that the cartridge spun into just the right position for firing. Assuming that you influenced the spins of neither the die nor the cylinder, neither the six nor the gunshot was intentional.'

'Then my shooting of you—how was that intentional?'

'Though your decision to play the shooting game was not a move in *that* game, it *was* one in what we might call the "life game", the possible outcomes of which are my life and my death. Since your decision to play the shooting game was a direct factor in my imminent demise, the fact of your shooting me would have to be considered intentional.

'Now, as to your *motives* . . .'

But before he can finish, Smith expires.

Swarthmore College   © DAVID ROSS 1978

INTENTIONAL ACTION, CHANCE AND CONTROL

By ERIC RUSSERT KRAEMER

IN ANALYSIS "Problem" No. 16 R. J. Butler asks us to compare the case of Brown throwing a six in an ordinary dice game with the case of Brown putting one live cartridge into an otherwise empty, six-chambered revolver, spinning the chamber, pulling the trigger while pointing at Smith, and killing Smith. Butler asks why it is the case that Brown does not intentionally throw the six whereas Brown does intentionally kill Smith if both situations have the same probability.

To answer Butler's question let us first consider both of his cases more carefully. In the case of the dice game there are three states of affairs to be noted:

(1) Brown's throwing a die

(2) Brown's throwing a six

(3) Brown's winning the dice game

Brown clearly does (1) intentionally, he does not do (2) intentionally, and, presumably, he does (3) intentionally. In the case of the shooting there are also three states of affairs to be considered:

(4) Brown's pulling the trigger

(5) Brown's firing the bullet

(6) Brown's killing Smith

Brown clearly does (4) intentionally, and he also does (6) intentionally. But, parallel to (2), he does not do (5) intentionally. Thus there really is no asymmetry between the two cases.

We may still wonder why Brown may do (1) and (4) as well as (3) and (6) intentionally but may not do (2) and (5) intentionally. The answer to this, I think, has to do with whether or not Brown has control over a particular situation. Brown presumably has control over whether he throws the die and over whether he pulls the trigger. Brown also has *some* control over whether he is the winner of the dice game in virtue of his having control over whether he plays the dice game or not. Further, Brown has some control over whether or not he kills Smith. For we assume that he has control over whether he pulls the trigger. But, although Brown has control over whether he throws a die he has no control over whether he throws a six. For he has no control over whether the die that he has thrown will turn up six. That he has control

over whether or not he will throw the die does not imply that he has some control over what the throw will be. So, Brown has no control over whether he will throw a *six*. Similarly, although Brown has control over whether he pulls the trigger and also has control over whether he spins the chamber, he has no control over whether the spun chamber will turn up a live cartridge in firing position. That Brown has control over whether he will pull the trigger which will release the contents of *some* chamber does not imply that he has control over releasing the contents of a specific chamber. So, he has no control over whether he fires the cartridge. Brown may desire that the six appear and that the cartridge be in firing position. But as he has no control over how the roll of the die or how the spin of the chamber turn out, he has no control over (2) or (5). Since Brown has no control over (2) or (5) he cannot do either of them intentionally.

Notice that our intuitions would be different with respect to these cases were we to discover that Brown had exercised psychokinetic powers over the rolling die or over the spinning chamber. For in such cases Brown would in fact have some control over these normally random events.[1]

NEITHER INTENTIONAL NOR UNINTENTIONAL

By E. J. Lowe

THE reason why, in the dice-game case, we would not say that Brown threw a six *intentionally* seems to be that he could not have known, when he threw the die, that it would certainly come to rest with the appropriate face uppermost. This suggests that we may analyse a proposition of the form 'A did X, not intentionally' as meaning something like 'A did Y, not knowing that X would certainly result, and X resulted'. (Y here may or may not be the same event or action as X: if it is the same, 'result(ed)' must be understood non-causally.) We must then distinguish three other closely related forms of proposition, as follows. 'A did X intentionally', which is analysed as 'A did Y, knowing that X would certainly result, and X resulted'. 'A did X unintentionally', which is analysed as 'A did Y, not knowing that X might possibly result, and X resulted'. And 'A did X, not unintentionally', which is analysed as 'A did Y, knowing that X might possibly result, and X

resulted'. According to these analyses, the following logical relation-
ships hold. 'A did X intentionally' entails (but is not entailed by) 'A did
X, not unintentionally'. 'A did X unintentionally' entails (but is not
entailed by) 'A did X, not intentionally'. 'A did X, not intentionally' is
consistent both with 'A did X unintentionally' and with 'A did X, not
unintentionally'. And 'A did X, not unintentionally' is consistent both
with 'A did X intentionally' and with 'A did X, not intentionally'.

    If the foregoing analyses are correct, it is true, in the dice-game case,
not only that Brown threw a six *not intentionally*, but also that he threw a
six *not unintentionally*. For while he did not know, when he threw the die,
that it would certainly come to rest with the appropriate face uppermost,
he did know that this might possibly happen. By parity of reasoning we
must similarly say, in the revolver case, that Brown killed Smith at once
not intentionally and *not unintentionally*. For on the one hand he did not
know, when he pulled the trigger, that Smith's death would certainly
result, while on the other he did know that it might possibly result. The
two cases are thus formally quite parallel. Why, then, are we inclined to
say in the dice-game case merely that Brown threw a six *not intentionally*
(ignoring the fact that he also threw a six *not unintentionally*) and in the
revolver case (incorrectly, as I would claim) that Brown killed Smith
intentionally? The answer lies in the distinctive *moral* features of the two
cases. In games of chance it is enough, to ensure fairness, that a player
should *not* be able to achieve a desired result *intentionally*: it would be
absurd to insist that he should only be able to achieve such a result
unintentionally. Therefore we tend to ignore as irrelevant the fact that
Brown also threw a six *not unintentionally*. It is quite otherwise in the
revolver case. A harmful act which is *not intentional* is not blameworthy
if it is also *unintentional*: but certainly is blameworthy if, as Brown's act
was, it is also *not unintentional*. However, most blameworthy acts, if
malicious (rather than merely callous), are straightforwardly *intentional*,
since the malefactor will normally use (unlike Brown) the surest means
to his evil ends. So we tend to regard Brown's clearly blameworthy and
malicious act as *intentional* instead of just *not unintentional*, despite the
contrary evidence.

A NOTE ON LOGICAL CONSTANTS

By THOMAS BJURLÖF

MUCH has been said about logical constants. In a recent article Christopher Peacocke suggested a criterion for an expression to be a logical constant relative to a (formal) language and a (formal) semantics or theory of truth.[1] As I will show, the criterion is too closely tied to the axioms of the truth theory.

Peacocke's criterion is framed in the Tarskian tradition. A theory of truth (or satisfaction) of this kind makes use of sequences of objects of the domain or universe of discourse. We can think of a sequence as a function that to each variable of a language L assigns an element of the domain D of L. If Seq is a sequence and v a variable, then $Seq(v)$ is an element of D. Two examples: the sequence Seq satisfies the formula $F(x)$ if and only if $Seq(x)$ is an F (to be precise: if $Seq(x)$ is an element of the set assigned to F by the interpretation of L), and $F(x)$ is true if and only if it is satisfied by all sequences; $(\exists x)\ F(x)$ is satisfied by Seq if and only if there is a sequence Seq_1 that differs from Seq at most in that $Seq(x) \neq Seq_1(x)$, and Seq_1 satisfies $F(x)$. Peacocke's idea can now be presented in the following way: $(\exists x)$ is a logical constant since to know whether a formula $(\exists x)\gamma$ is true we only need to know what sequences satisfy γ and the satisfaction condition for existentially generalized formulas, but nothing about specific elements of D. What makes '$(\exists x)$' a logical constant is then the fact that knowledge of properties and relations of elements of the domain is not required for us to know whether $(\exists x)\gamma$ is true or false, given knowledge as to what sequences satisfy γ. Peacocke's statement of the criterion in its generality is the following:

> α is a logical constant if α is noncomplex and, where the syntactic category of α is $\tau/\sigma_1, \ldots, \sigma_n$, for any expressions $\beta_1, \ldots, \beta_n$ of categories $\sigma_1, \ldots, \sigma_n$, respectively, given knowledge of
> (a) which sequences satisfy those β_i which have satisfaction conditions, and of
> (b) which object each sequence assigns to those β_i which are input to the assignment function, and of
> (c) the satisfication condition or assignment clause for expressions of the form $\alpha(\gamma_1, \ldots, \gamma_n)$
> one can know a priori which sequences satisfy the expression $\alpha(\beta_1, \ldots, \beta_n)$ of category τ, or which object any given sequence assigns to $\alpha(\beta_1, \ldots, \beta_n)$, in particular without knowing the properties and relations of objects in the sequences.[2]

[1] C. Peacocke, 'What is a Logical Constant?' *The Journal of Philosophy*, Volume LXXIII, Number 9, 6 May 1976.
[2] C. Peacocke, *ibid.*, pp. 225–6.

To show that Peacocke is mistaken I will make use of Hilbert's ε-operator. In semantic terms, the ε-operator picks a "representative" from a class of elements. Assume that we have a first-order theory with the epsilon axiom:

(1) $A(t) \rightarrow A(\varepsilon_x A(x))$.

Intuitively, this axiom says that if something t is an A, then the representative of A is also an A. Given (1) we can prove the following formula (the proof is quite simple, for details see *Hilbert and Bernays*[3]):

(2) $(\exists x) A(x) \longleftrightarrow A(\varepsilon_x A(x))$.

(2) can be thought of as a contextual definition of the existential quantifier. The universal quantifier is defined in terms of $\neg(\exists x)\neg$. Assume that the theory also contains the following axiom schema:

(3) $A(t) \rightarrow \varepsilon_x A(x) \leqslant . t$

If the domain is the set of natural numbers we may let $\leqslant$ be the usual ordering on the natural numbers. A model for the theory would include a function f such that $f(A(x))$ is the least x for which $A(x)$ if A is assigned a non-empty set, otherwise an arbitrary element of the domain, say o. $\varepsilon_x A(x)$ is assigned $f(A(x))$. (For a theory of this kind see *Tait*.[4]) A sequence *Seq* satisfies a formula $A(\varepsilon_x A(x))$ if and only if $f(A(x))$ is an element of the set assigned to A. The satisfaction condition for an existential generalization can then, following (2), be defined as the satisfaction condition for the equivalent epsilon formula.

(*Hilbert and Bernays* do not have (3). To simplify matters I assumed that we dealt with the natural numbers. This assures that each non-empty set has a least element. My arguments would go through equally well without these assumptions. For a discussion of Hilbert's and Bernays' treatment of epsilon-terms, where the representative of a non-empty set is an element of the set but there is no way to find out which, see *Carnap*[5]. On how to develop the semantics of this theory, see *Asser*[6].)

For a theory of the kind sketched above it follows that to find out whether a sequence satisfies an existential generalization (and hence a universal generalization) we need to have knowledge of properties and relations of specific elements of the domain: a sequence *Seq* satisfies

[3] Hilbert and Bernays, *Grundlagen der Mathematik II*, Zweite Auflage, Springer Verlag 1970; pp. 9–18.
[4] W. W. Tait, 'The Substitution Method', *The Journal of Symbolic Logic*, Volume 30, Number 2, June 1965.
[5] R. Carnap, 'On the Use of Hilbert's ε-operator in Scientific Theories', in *Essays on the Foundations of Mathematics Dedicated to A. A. Fraenkel on His Seventieth Anniversary*, pp. 156–64, Jerusalem 1961.
[6] G. Asser, 'Theorie der logischen Auswahlfunktionen', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, Volume 3, 1957.

$(\exists x)A(x)$ if and only if $\varepsilon_x A(x)$, the representative of A, is an A. (Notice that the exact character of Seq is irrelevant.) But if so $(\exists x)$ is in this theory not a logical constant according to Peacocke's criterion!

It would be incorrect to say that the existential quantifier above is not the ordinary existential quantifier since a formula $(\exists x)A(x)$ is equivalent to an epsilon formula whose epsilon term is assigned a specific object of the domain, and hence $(\exists x)A(x)$ ought to be read as a sentence about this object. The error is similar to assuming that since in a finite domain a universal generalization is equivalent to the conjunction of its instances it does not differ in meaning from the conjunction. It would further be arbitrary to defend Peacocke's criterion by saying that $(\exists x)$ above is complex. If existential generalizations had been given the usual satisfaction condition, $(\exists x)$ would have been a logical constant on Peacocke's account, but how could this change of satisfaction or truth conditions determine whether $(\exists x)$ is complex or non-complex?

There is a way to defend the criterion that is not open to Peacocke. Although, in the theory we gave, to know whether a sequence Seq satisfies a formula $(\exists x)A(x)$ according to the satisfaction conditions (or in Peacocke's terms: the axioms of the truth theory) we need to have knowledge about $f(A(x))$, namely whether it is an element of the set assigned to A, there is an equivalent satisfaction condition that does not require knowledge of this kind. This is, of course, the ordinary satisfaction condition for existential generalizations. That Peacocke is not allowed this move can be seen from the following quote:

> If one accepts the Hilbert-Bernays definition of identity by exhaustion of lexicon which Quine advocates, identity will not be a logical constant on the criterion simply because it will not be a primitive expression that is handled by some axiom of the truth theory of the language.[7]

This is more than we need, since it excludes all expressions that do not have a truth axiom associated with them from being logical constants. If $(\exists x)$ is defined in terms of $\neg(\forall x)\neg$ for example, and $(\forall x)$ is associated with a satisfaction condition, or truth axiom, but not $(\exists x)$, then $(\exists x)$ is not a logical constant in this language.

One reason to think of $(\forall x)$ and $(\exists x)$ as logical constants is that formulas $(\forall x)A(x)$ and $(\exists x)A(x)$ do not say anything about particular objects of the domain. If Peacocke tried to capture this he erred in identifying what a formula including a logical constant says with its condition of satisfaction in the axioms of the truth theory.[8]

University of North Carolina at Chapel Hill   © Thomas Bjurlöf
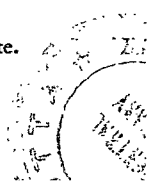
A REPLY TO BJURLÖF'S OBJECTION

By CHRISTOPHER PEACOCKE

IF we have a criterion of logical constanthood that applies only to primitive expressions that are assigned semantic properties directly by the axioms of a truth theory, it seems reasonable to extend the criterion by deciding to call a defined expression a logical constant if all the primitive expressions in its definition are logical constants by that original criterion. So if '∃' is defined in terms of the ε-operator, in order to see whether '∃' so defined is a logical constant we must apply the original criterion to see whether the ε-operator is a logical constant.

In his arithmetical example, Bjurlöf in effect defines the ε-operator as follows, where 's' ranges over sequences of natural numbers, 'x' over the natural numbers themselves, 'v' over variables of the object language, '$s(v/x)$' abbreviates 'the sequence that differs from s at most in assigning x to v', and "*" is the assignment functor:

$$(1) \quad s^*(\ulcorner \varepsilon v A(v) \urcorner) = \begin{cases} \mu x(\mathrm{sats}(s(v/x), \ulcorner A(v) \urcorner) \text{if} \\ \quad \exists x(\mathrm{sats}(s(v/x), \ulcorner A(v) \urcorner)) \\ O \text{ otherwise.} \end{cases}$$

To apply the criterion of logical constanthood, we ask: is it the case that given a knowledge of which sequences of natural numbers satisfy $\ulcorner A(v) \urcorner$ together with a knowledge of the axiom (1) one can infer *a priori* which object a given sequence assigns to $\ulcorner \varepsilon v A(v) \urcorner$? Yes, one can: of these sequences that one knows to satisfy $\ulcorner A(v) \urcorner$ one takes the least number such that one of these sequences assigns that number to the variable v, and O if there are no such sequences. Here we have made use of the fact that 'less than' holds *a priori* between a given pair of numbers if it holds at all. (This strategy need not make all arithmetical functors logical constants: a restriction may be adopted to *a priori* relations failure to recognize the holding of which casts doubt on grasp of the notions in question.) So far, then, we can say that if '∃' is defined in terms of such an ε-operator it will be a logical constant too.

That was *all* we had to show in order to demonstrate that such an ε-operator is a logical constant by the original criterion. When Bjurlöf writes that my idea is that ' "(∃x)" is a logical constant since to know whether a formula $\ulcorner (\exists x) \gamma \urcorner$ is true we only need to know what sequences satisfy γ and the satisfaction condition for existentially generalized formulas, but nothing about specific elements of D', that occurrence of 'nothing' should be replaced with the phrase: 'nothing not entailed by the given information about which sequences satisfy γ and the given satisfaction condition'. The phrase 'in particular without knowing the

properties and relations of objects in the sequences' that comes at the end of the general statement of my criterion that Bjurlöf quotes is a gloss on what precedes it, and not an additional requirement: if it were, then no amount of knowledge of identities of objects in the sequences would allow even the definite description operator to count as a logical constant.

Have I relied too heavily upon the chosen properties of a special case, in which there is an *a priori* wellordering by the 'less than' relation of all the elements of the domain? Let us then consider a case without this property, in which the range of quantification is all persons (in this or any other galaxy). It is to be noted incidentally that there is an onus too upon the supporter of Bjurlöf's objection to develop this case, for the claim of any expression to be a logical constant must be weakened if it is by my criterion a logical constant with respect to a truth theory for an object language with a given range of quantification, but either fails the criterion or is unintelligible for a truth theory for an otherwise similar object language having a wider range of quantification.

For the more general case in which there is no *a priori* wellordering of the domain, Bjurlöf refers us to Asser's model theoretic semantics for languages containing the ε-operator.[1] In these semantics, specification of a model involves specification of a choice function, and (relative to a given sequence s) in a given model $\mathcal{M}$, $\ulcorner εvA(v) \urcorner$ denotes

$$\varphi(\{x : \text{sats}(s(v/x), \mathcal{M}, \ulcorner A(v) \urcorner)\})$$

where φ is the choice function of the model. Now since my criterion is concerned with absolute truth theories, and not with truth in a model, if Asser's semantics is to be used for the case of quantification over all persons, then we must specify a particular choice function in the truth theory for an object language. How are we to specify it?

We cannot specify it by enumeration. Not only is that not feasible, but even if it were it would be theoretically incorrect to do so. For the specification of the function will occur on the right hand side of T-sentences of the theory—that is, in specifications of what object language sentences *say*—and it is absurd to maintain that speakers must know such an enumeration to understand the language. The other alternative is to specify the function by some property which, given a set of persons, picks out one person from that set (e.g. the firstborn). But then 'ε' would not be a logical constant, either intuitively or by my criterion: given a knowledge of which sequences satisfy $\ulcorner A(v) \urcorner$, to know which object $\ulcorner ε vA(v) \urcorner$ denotes (relative to a given sequence) one would need the further *a posteriori* information to the effect that a certain person in a given set is the one with the property used in the specification of

---

[1] There is a description in English of Asser's semantics in A. C. Leisenring's expository book *Mathematical Logic and Hilbert's ε-Symbol* (London: McDonald, 1969).

the choice function. Nor could we avoid problems by existentially quantifying over choice functions and saying in the truth theory only that there is such a function fulfilling the assignment axiom for 'ε': for then the truth theory would fail to say explicitly what the truth conditions of object language sentences are.

What of a homophonic theory? If the role of 'ε' in the object language is not determinate beyond what is fixed by the requirements that all instances of '$A(t) \supset A(\varepsilon v A(v))$' and '$\forall x(A \equiv B) \supset \varepsilon x A(x) = \varepsilon x B(x)$' come out true, then it is clear that there will be many sentences of the object language containing 'ε' for which these requirements are not sufficient to determine truth conditions. This results in an obstacle to giving a homophonic theory, for one of the sentences whose truth conditions are left unfixed in a metalanguage containing the object language is

$$ s^*(\ulcorner \varepsilon v A(v) \urcorner = \varepsilon x \text{sats}(s(v/x), \ulcorner A(v) \urcorner) $$

the natural axiom for a homophonic theory. (These difficulties seem ultimately to flow from the attempt to treat an expression introduced to play the role of a quantifier as a genuine term for an object.)

The upshot seems to be this. '∃' defined in terms of 'ε' may be a logical constant relative to a truth theory for an object language with an ontology of natural numbers; but equally my criterion pronounces it as such. On a narrower notion of logical constanthood, one might require that an expression be a logical constant in the original sense with respect to *all* truth theories that differ from one another only in the richness of the ontology of the object language. In that sense, Hilbert's ε-operator is not plausibly a logical constant, nor do I make it one, nor is '∃' so generally definable in terms of 'ε'.

All Souls College, Oxford   © CHRISTOPHER PEACOCKE 1978

KILLING, LETTING DIE, AND JUSTICE

By RICHARD O'NEIL

DANIEL DINELLO in his paper 'On Killing and Letting Die' (ANALYSIS, 31.3) argues that the killing/letting die distinction can have moral significance at least in situations in which one must choose between the lives of two persons. He presents the following example:

> Jones and Smith are in a hospital. Jones cannot live longer than two hours unless he gets a heart transplant. Smith, who had had one kidney removed, is dying of an infection in the other kidney. If he does not get a kidney transplant, he will die in about four hours. When Jones dies, his one good kidney can be transplanted to Smith, or Smith could be killed and his heart transplanted to Jones. Circumstances are such that there are no other hearts or kidneys available within the time necessary to save either one. Further, the consequences of either alternative are approximately equivalent, that is, heart transplants have been perfected, both have a wife and no children, etc. (pp. 85-6)

Dinello points out that, intuitively, it clearly would be wrong to kill Smith and save Jones, rather than letting Jones die and saving Smith. This, he concludes, must be because the killing/letting die distinction itself makes a moral difference since the consequences in this case are equivalent.

But this conclusion is incorrect. What makes killing Smith wrong in this case is not the mere fact that we would be killing him as opposed to letting Jones die. To see this, suppose a physician transplanted Smith's heart to Jones but kept Smith alive until his kidney failed by placing him on an artificial heart machine. By doing so the physician would be letting Smith die (of his kidney infection) rather than killing him. Yet it seems just as objectionable to save Jones in this way as to save Jones by killing Smith. If so, Dinello is mistaken in thinking his example demonstrates the moral significance of the killing/letting die distinction.

Why is it wrong to save Jones at the cost of killing Smith or letting him die? I believe it is wrong because it is an instance of distributive injustice. There exists a random procedure for distributing the evil of death the fairness of which we have come to accept: the 'natural lottery' of fate. This is not so say we consider it wrong to interfere with nature's processes in order to save lives. It is only that when we are at the point when we must choose between one person's death and another's we find that there already exists the randomization involved in letting the loss lie where it falls. To introduce another lottery or some other method for choosing at this point is unfair to the winner of the initial 'lottery'.

The time at which Smith and Jones are fated to die after equal efforts have been made to save them is a sufficiently random circumstance to enable us to say that they have had an equal opportunity to gain by the other's death. If then Jones claims that we should interrupt the workings of fate by killing Smith or by removing Smith's heart before Smith dies, he is suggesting that we nullify a fair procedure for implementing distributive justice. It is this which makes the killing of Smith wrong in the situation Dinello describes, not the fact that we would be killing Smith versus letting Jones die.

Emory and Henry College   © RICHARD O'NEIL 1978

LEWY ON C. I. LEWIS AND ENTAILMENT

By Peter M. Simons

DR. C. LEWY has claimed[1] that the only ground one can have for rejecting the rule of disjunctive syllogism, 'from $-A$ and $A \vee B$ infer B', as an acceptable principle of entailment, is rejection of the principle of bivalence, so that those who accept bivalence but reject disjunctive syllogism have no grounds for this latter rejection. I wish to suggest that this claim is ill-founded.

C. I. Lewis showed[2] that anyone prepared to accept certain principles of inference as characterizing entailment would also have to accept the following paradoxes:

(1) $(A \& -A) \rightarrow B$
(2) $A \rightarrow (B \vee -B)$

(writing '$\rightarrow$' for 'entails'.)

Lewis's proofs—call them Proof (1) and Proof (2) respectively,[3] rest on principles of inference of which, in each case, as Lewy shows,[4] suspicion can fall on one alone; in Proof (1) the principle of disjunctive syllogism

(a) from $-A \& (A \vee B)$ infer B

and in Proof (2)

(b) from A infer $(A \& B) \vee (A \& -B)$.

Lewy contends that the suspicious inference appealing to (b) uses a *suppressed* premiss; so to make this premiss explicit we must substitute rule (c):

(c) from $A \& (B \vee -B)$ infer $(A \& B) \vee (A \& -B)$

though then the paradox also disappears, since with (c) we cannot prove the paradox (2), only the weaker

(3) $(A \& (B \vee -B)) \rightarrow (B \vee -B)$

which is a perfectly acceptable principle of entailment, an instance of conjunction elimination. Lewy further points out that the formula
$$p \supset ((p \& q) \vee (p \& -q))$$
which is closely related to (b), is not a tautology (i.e. does not always

[1] Lewy, *Meaning and Modality*, Cambridge, 1976, pp. 108–117.
[2] Lewis and Langford, *Symbolic Logic*, New York, 1932, pp. 250–1.
[3] These are well-known and are not repeated here.
[4] Lewy, *op. cit.*, p. 113.

take the designated value 1) in Łukasiewicz's 3-valued propositional calculus, whereas the formula similarly related to (*c*), namely

$$(p \ \& \ (q \lor -q)) \supset ((p \ \& \ q) \lor (p \ \& \ -q))$$

is a 3-valued tautology.

Since the Proof (2) is repaired, but also rendered ineffectual, by substitution of (*c*) for (*b*). Lewy argues that one intuitively expects the same thing to work for (*a*) in Proof (1). He accordingly alters the suspicious rule (*a*) to

(*d*) from $-(A \ \& \ -A) \ \& \ -A \ \& \ (A \lor B)$ infer B

but shows firstly that this sanctions the paradox

(4) $((A \ \& \ -A) \ \& \ -(A \ \& \ -A)) \to B$

and secondly that both

$$(-p \ \& \ (p \lor q)) \supset q$$

and

$$(-(p \ \& \ -p) \ \& \ -p \ \& \ (p \lor q)) \supset q$$

fail to be 3-valued tautologies.

He concludes that Proofs (1) and (2) are significantly different from one another. Proof (2) uses (*b*), which suppresses a premiss; restoring this renders the rules used acceptable, but innocuous, since the paradox (2) can no longer be proved; only the harmless (3). But a similar attempt at restoration in the case of (*a*) does not succeed, so what is here "suppressed" is no premiss but the metalogical principle of bivalence. Accordingly, Lewy concludes, the only reason for rejecting (*a*) and with it Proof (1) must be rejection of bivalence.

Opponents of (*a*), notably Anderson and Belnap, have not taken this line. They argue that (*a*) cannot work for purely extensional senses of 'or', whereas any sense for which it will work invalidates the rules of disjunction introduction, 'from A infer $A \lor B$, from B infer $A \lor B$', which Lewis also used in Proof (1), so that the proof rests on an equivocation.[5] This is their answer to Lewy's plea for an explanation of the unacceptability of (*a*),[6] whereas Lewy claims that they 'cannot get round that proof in the way they attempt',[7] by rejecting (*a*), since they accept bivalence and, according to Lewy, (*a*) and bivalence stand or fall together.

Lewy is unjustified in this conclusion. His argument depends, as we have seen, on taking Proofs (1) and (2) to be significantly similar in some way, so that we intuitively expect them to go astray in similar ways. He then claims to expose a lack of similiarity between them.

---

[5] Anderson and Belnap, *Entailment*, Vol. 1, Princeton, 1975, pp. 163–7.
[6] Lewy, *op. cit.*, p. 116.
[7] *Ibid.*

The similarity resides in the fact that the formulae in (1) and (2) are *dual*, a fact which Lewy does not explicitly recognize. Two propositional formulae are dual to one another when their logical behaviour is alike except for a thorough interchange of the logical roles of truth and falsity. So A & B and $A \lor B$ are dual because the former is *true* if and only if both A and B are *true*, whereas the latter is *false* if and only if both A and B are *false*. Propositional variables and their negates are self-dual, and A implies B if and only if the dual of B implies the dual of A.[8]

If the formulae in (1) and (2) are dual, as they are, then it makes sense to suppose that the proofs of (1) and (2) are suspicious in dual ways, and hence that the suspicious principles (*a*) and (*b*) can be disarmed in dual ways. Some such thinking lies behind Lewy's attempt at repairing them both, but his attempt fails because he does not repair (*a*) and (*b*) in dual ways. He restores a lost premiss to both (*a*) and (*b*); this works for (*b*) but not for (*a*). But to restore a lost premiss is to restore a lost *conjunct* in the *antecedent* of a conditional; the dual to this is restoring a lost *disjunct* in the *consequent*.

So if we repair (*a*) dually to the way Lewy successfully repaired (*b*) we get the perfectly acceptable

(*e*) from $-A$ & $(A \lor B)$ infer $B \lor -(A$ & $-A)$

which blocks the proof of (1); we can now only prove the innocuous

$$(A \text{ \& } -A) \rightarrow (B \lor -(A \text{ \& } -A))$$

which does not depend on bivalence, since

$$(-p \text{ \& } (p \lor q)) \supset (q \lor -(p \text{ \& } -p))$$

is a 3-valued tautology.

Furthermore, if we take the dual to Lewy's own unsuccessful way of trying to repair (*a*) in the case of (*b*), we get the rule

(*f*) from A infer $(A$ & $B) \lor (A$ & $-B) \lor (B \lor -B)$

and not only is

$$p \supset ((p \text{ \& } q) \lor (p \text{ \& } -q) \lor (q \lor -q))$$

not a 3-valued tautology, taking the value O for $p=1$, $q=\frac{1}{2}$, but rule (*f*) leads, by an argument like Lewy's for (4), to the paradox

$$A \rightarrow ((B \lor -B) \lor -(B \lor -B)).$$

Hence Lewy's attempt to show that Proofs (1) and (2) are significantly different breaks down. (1) and (2) are dual, but (*c*) and (*d*) are not. The dual of (*c*) is (*e*), which allows bivalence to be retained without paradox,

---

[8] See e.g. Quine, *Methods of Logic*, 3rd edition, London, 1974, Section 12.

whereas the dual of (*d*) is (*f*), which, like (*d*), leads to paradox. So (*a*) depends on bivalence no more and no less than (*b*). It might at first be supposed that the only form of suppression worth the name is suppression of a premiss. But suppression of a conclusion, in this case disjoined, is no less suppression. If we omit the conjunct − *A* in the premiss of the rule 'from *A* & − *A* infer − *A*' the result is exactly as if we had omitted the disjunct *A* from the conclusion of the rule 'from *A* infer *A* ∨ − *A*'; in each case we get the outrageous 'from *A* infer − *A*'.

Whether in ordinary discourse we omit conjoined premisses more often than disjoined conclusions is an empirical question. If we do, this may explain why Lewy moved to retain the "similarity" as he did, but the question is beside the point in the present context.

Bolton Institute of Technology   © PETER M. SIMONS 1978

ACQUIRING AND POSSESSING KNOWLEDGE

By ALAN R. WHITE

IT is sometimes suggested[1] that Gilbert Ryle considered 'know' an *achievement* or *got it* verb, that is as signifying not only that some performance has been gone through, but also that something has been brought off by the agent going through it (*Concept of Mind*, 149–153). And certainly much of what Ryle says about knowing something does imply or even express this view. First, he groups (pp. 130, 152–3, 239) 'know' with such verbs as 'find', 'discover', 'prove', 'solve', 'see', 'detect', 'deduce', 'recall', 'conclude', all of which latter he rightly holds are achievements or, perhaps better, acquisitions.

Secondly, he sometimes (e.g. 150–1) moves from the view that achievement verbs 'signify not only that some performance has been gone through but also that something has been brought off by the agent going through it' to the view that 'in applying an achievement verb we are asserting that some state of affairs obtains over and above that which consists in the performance, if any, of the subservient task activity'.

[1] e.g. I. Scheffler, *Conditions of Knowledge* (1965) pp. 28–31, who—despite criticisms of a different kind from mine of Ryle—accepts the thesis; cp. E. M. Adams 'On knowing that', *Philosophical Quarterly* 8 (1958) pp. 300–06, who says that 'know' is both a 'capacity dispositional achievement verb' and an 'episodic achievement verb'. W. H. F. Barnes in 'On seeing and hearing' *Contemporary British Philosophy*, III, 67–9 (edited H. D. Lewis) does distinguish for perceptual verbs, like 'see', 'hear', etc. an 'achievement' use and a 'state' or 'experience' use, but not for 'know'. I think the distinction is implausible for perceptual verbs.

Thirdly, he sometimes groups 'being right', which is a characteristic of knowing, that is, the possession of knowledge, with 'succeeding', which is not. Fourthly, he overlooks (p. 239, 134) the fact that the question 'How do you know?' can be used either, and perhaps more commonly, to ask how someone got hold of his knowledge or to ask him how it is that he possesses that knowledge. And, hence, fifthly, much of his discussion (e.g. pp. 40-1) of how one acquires either knowledge of how to do something or the ability to do it does not distinguish it from the possession of that know-how, or that ability.

In so far as Ryle does undoubtedly say these things, he is wrong. He has rightly distinguished both 'discover', 'solve', 'detect', etc. and knowing from searching, puzzling over, looking for, etc.; but wrongly concluded that 'know' is to be contrasted in the same way with these as are 'detect', 'solve', 'discover'.

To know is not to achieve something, it is to possess it, even if—which is doubtful—the verb 'know' in English is sometimes used as an abbreviation for 'get to know' as when one asks someone when did he first know that the butler was an ex-convict. Similarly, we can say both that someone 'got' sixpence on his birthday and that he has now 'got' sixpence. Yet all of the above arguments assimilate knowing, that is possessing knowledge, to achieving or acquiring knowledge. 'Know' is not in fact in the same class as 'discover', 'prove', 'solve', 'detect', 'deduce', or 'conclude'. Though one can get to know something, as one can discover, prove or solve it, at a particular point in time, one knows something for or during a particular time. One can know something for years, but not get to know, solve or discover it for years. On the other hand, whereas one can take years to solve, discover or get to know something, one cannot take years to know it. One can forget how to discover, prove, or solve something, but not forget how to know it; though one can, of course, forget the thing itself. 'Knowing', unlike 'discovering', 'proving' and 'solving' does not signify 'that something has been brought off', but that 'some state of affairs obtains over and above' either a corresponding task or a corresponding achievement. 'Succeeding' is not 'being right', though success in certain tasks may result in being right. One can know well or better, thoroughly, off by heart, etc., but not discover, prove, detect or come to know in these ways. One can have the ability to solve, discover or prove something, but not the ability to know it. Abilities, but not achievements, can be built up by practice. And one can put into practice one's abilities or one's knowledge but not one's achievements. 'Knowing', i.e. 'having knowledge', is no more the same as coming to know, i.e. achieving or acquiring knowledge, than 'becoming President' is the same as 'being President' or 'arriving at the airport' is the same as 'being at the airport'.

Ryle's real or apparent adherence to an achievement view of the use

of the verb 'know' is, moreover, inconsistent with much of what he says—and, in my opinion, rightly says—about knowing in other parts of, e.g., *The Concept of Mind.* For there (e.g. 44–6, 59, 133) he undoubtedly argues, not that 'know' signifies an achievement, but that it signifies a possession, namely the possession of a disposition, or, rather, an ability or a capacity. Significantly, the index to the *Concept of Mind* refers under 'know' not to the idea of an achievement, but to those of a disposition and a capacity. 'Know' is said to signify a capacity to 'bring things off' or 'get them right', not simply 'bringing them off' or 'getting them right' themselves. Nor is a capacity or disposition to achieve a type of achievement[2], e.g. 'a capacity achievement', any more than a capacity or disposition to act or suffer is a type of act or a type of suffering. To prove or to detect are achievements, whereas the ability or capacity to prove or to detect, though they may be capacities to achieve, are not themselves achievements; yet it is with 'prove' and 'detect' that Ryle earlier compared 'know'. Certain tasks may enable one to achieve what one is after, whether it is knowledge or some other prize, but what one does achieve is not itself an achievement. 'Knowing' may presuppose 'having learnt', but it is not equivalent to it (cp. 226). Nor is the distinction between an achievement and a possession confined by Ryle to knowing how to do something, for he also says that knowing that something is so is to have and retain a store house of truths, not that it is the acquisition of such a store house. Furthermore, Ryle himself emphasises in another context (pp. 301–3) the difference between, e.g., having a plan, an argument or a proof and getting any of these; and in his *Dilemmas* (p. 109), he recognises that, e.g., 'discovering' is 'coming to know' and not 'knowing', even though one obviously knows that which one has come to know.

Hence, I conclude, first, that Ryle wavered between viewing 'know' as a verb expressing the achievement of something and viewing it as a verb expressing the possession of it; and, secondly, that the former view is mistaken.

University of Hull © ALAN R. WHITE 1978

[2] *pace* Adams, *op. cit.*

A THEODICY

By JOHN D. McHARRY

THE theological problem of evil is, in stark terms, to refute the following argument:

(1) God exists and is all good, all knowing, and all powerful.

(2) If God exists, the world must be in accord with His wishes, i.e., the best of all possible worlds.

(3) Therefore, this *is* the best of all possible worlds.

(4) This is *not* the best of all possible worlds; it contains too much evil.

(5) Therefore, (1) is false; there is no God.

Setting aside those replies which would give up one or another of the divine attributes, most traditional replies would seem to involve denying (4). They would do this by adducing reasons for holding that this is, indeed, the best of all possible worlds. Familiar among such replies are those which would attempt to show that all evils are merely apparent, disappearing in some global harmony, and those which attempt to show that what evils there are are all necessary conditions for a greater amount of good, such as the free will defence. Needless to say, none of these replies has been totally convincing.

Both the initial argument and the attempted refutations, however, seem to accept a covert premise that there is and can be but one actual world among all those which are possible. Given certain assumptions, this need not be the case.

At this point one might object that given two or more distinct "worlds", *the* World, or God's World, just *is* the collection of both or all of them. There are two ways of responding to this. The first is to clarify what is meant by a "world"; perhaps the collection of all things spatiotemporally related to one another. On this interpretation, two such disjoint collections will not constitute one world. The second response is to accept the criticism and to hold that the classical error lies in holding that the world is and can be constituted of only one such collection. It may in fact consist of more than one.

I shall for present purposes construe a world as a collection of spatiotemporally interrelated objects. Whether such a world could be a mere part of a more inclusive world I shall not decide. Interpreting 'world' in this way is both adequate to characterize the classical discussion and to demonstrate where it went wrong.

Nor is my notion of a world unprecedented, although it is clearly *not* that used in the analysis of counterfactuals. My notion is similar to one of Bondi's: everything in an observer's past and future light cones

(*Cosmology*, Cambridge 1952, p. 10). The set of all the possible worlds of which I speak may indeed be the set of all cosmological solutions to the gravitational equations, as Gödel has suggested (*Albert Einstein: Philosopher-Scientist*, ed. P. A. Schilpp, Evanston, 1949, p. 559).

I shall now show how it is that the possibility of more than one actual world allows a refutation of the argument given in the first paragraph.

Premise (2) of that argument is to be interpreted as expressing the position that what God creates or countenances must be the best that could be. Furthermore, if *this* world is the *only* world, it must therefore be the best of all possible worlds. However, if there could be more than one actual world, this one need be best only if there could be as many optimal worlds as there are worlds, or if God can only create that world or those worlds which *are* optimal. Given certain assumptions, neither condition holds.

Say God creates the best of all possible worlds, whatever that is. Now, if there can be more than one Space, it would seem He can create more than one world. Let us assume therefore, that there *can* be more than one Space. (cf. Quinton, 'Spaces and Times', *Philosophy*, 1962.)

Now there can be more than one world, but we have as yet no reason to suppose that God could not just create as many copies as He would of the best of all possible worlds. Indeed, if Spaces were like pigeon holes, this is just what He would do, given some principle of plenitude.

We need, therefore, to assume a strong version of Leibniz' law of the identity of indiscernibles: there cannot be two distinct entities which do not differ in respect to at least one property. Given this, there cannot be undifferentiated Spaces lined up like pigeon holes, nor can there be multiple copies of the best of all possible worlds. Each world must differ from all others in some way, however trivial.

If there can be but one best of all possible worlds, and God has created more than one world, it is easy to see why *this* one need not be optimal. If, however, there are variations on some optimal design which make no difference in the value of a world, such as making everything a bit larger or smaller, then there might be a rather large, perhaps infinite, class of optimal worlds. If this class is, nevertheless, a proper subset of the class of actual worlds, the problem is resolved.

I must now show how it is that God would actualize more than the best of all possible worlds or the class of optimal worlds.

I am somewhat embarrassed that I must concede I don't know why He would actualize any world at all. It has been claimed that God creates the best of all possible worlds because it is the best, and perhaps this is as close as we can come to an explanation. If this *is* the reason, then we have an account for all members of the class of optimal worlds. Consider the situation where only these have been created. So long as

the next best world or class of worlds is better than nothing, the best that God can do given the restriction that He cannot make exact copies of existing worlds is to actualize these worlds. (If the set of such classes is dense, this statement is inadequate. So long as the set can be ordered, however, the statement can be made more rigorous, if less intelligible.) This argument can of course be applied recursively to show why He would actualize *every* world which is better than nothing. (I admit that there may be some possible worlds which, being on the whole of negative value, would not be actualized.)

A rather difficult objection can be raised at this point. Consider a composer who has written his finest work. Is it better if he goes on to produce something not as good, or if he just stops? In other words, is it better to create the best plus something less good, or to create the best only?

If we accept the analogy, it seems we can find cases which go both ways. Perhaps it is better not to copy a tune except for adding a sour note at the end. On the other hand, is it not better to go ahead and write a completely *different* tune which, although not quite as good as the first, is nonetheless itself a thing of beauty?

The analogy, however, is somewhat strained. It is hard to believe that the value of a world derives from the originality of the creative work, or that its value is as upset by occasional "dissonances" as is that of a tune. While this objection raises certain qualms, it is on the whole not damning.

Although I have described a situation such that *if* it obtained the problem of evil would dissolve, it is fair to ask what possible evidence one could have that such a state of affairs is in fact the case. Clearly we could have no communication with any other world; if we did it would simply be a remote part of our world and of no use in my argument.

I would note that the demand that we have evidence of such worlds, or that we have the possibility of such evidence, rests on the notion that meaningfulness is tied to the possibility of verification. While this may in fact be true, it is an improper objection to *this* argument. A theodicy is, after all, an attempt to defend theism against one, very specific, objection. To bring in the question of verification is simply to raise *another* objection to the theistic theory, which is part and parcel unverifiable.

I hold, nevertheless, that the existence of the other worlds of which I speak is, in a sense, indirectly verifiable. Given this less than optimal world and the requirement that theism must, in order to be viable, be consistent with its lack of perfection, then if my theodicy is the only valid one, any evidence for the existence of God is at the same time evidence for the existence of the other actual worlds which I propose.

I.B.M. Corporation, Hopewell Junction, New York

CHANDLER ON CONTINGENT IDENTITY

By JOHN L. KING

HUGH S. CHANDLER[1] has presented an alleged counter-example to Saul A. Kripke's principle[2] that proper names are rigid designators, and hence to the principle that identity statements using proper names as designators are non-contingent. Chandler's argument rests, however, on an assumption which is left tacit and undefended in his paper. In the present note I shall identify this assumption and argue that a proponent of Kripke's principle need not accept it.

Chandler begins with Hobbes' story (S1) of a ship (a) whose parts are replaced one by one until the ship (c) which stands in its place consists entirely of replacement parts. Since a and c satisfy the familiar spatiotemporal continuity criterion, one is inclined to identify them. But a's cast off parts are collected, in the story, by a man who uses them to build a ship (b) exactly like a. Since a and b consist of the same parts, assembled according to the same plan, one is inclined to identify them. Since b and c are numerically distinct ships, however, at most one of the identities '$a=b$' and '$a=c$' is true.

Chandler maintains that spatiotemporal continuity is the "dominant" relationship in cases of this type, so that $a=c\neq b$. But he attempts to show that the non-identity of a and b is contingent by relating a story of his own (S2):

> Suppose that a's planks had been removed one by one *without being replaced*. b is then constructed just as in Hobbes' story. In this case a and b are the same ship. What it comes to is that Theseus' ship is transported from one place to another by being disassembled and then reassembled.[3]

If a and b are distinct in S1 but numerically the same in S2, then we have examples of contingent non-identity and contingent identity. Further, if we give the ships proper names, b's name will be a non-rigid designator of a, since it will designate a in S2 (where $a=b$) but not in S1. Finally, by substituting the proper names for 'a' and 'b' in '$a=b$', we can generate a contingent identity statement using proper names as designators. The principles mentioned above thus appear to be refuted.

Let us grant that the ship constructed in S1 from a's *replaced* parts is not a, and that the ship constructed in S2 from a's *disassembled* parts is a. Nothing of interest follows unless we also assume that the former and the latter are one and the same ship. Chandler makes this assumption implicitly and without defence when he says in S2 that b is constructed 'just as in Hobbes' story'. Now some ship or other, say, b^*, is constructed in

[1] 'Rigid Designation', *Journal of Philosophy* LXXII, (July 17, 1975), pp. 363–369.
[2] 'Identity and Necessity', in Milton K. Munitz, ed., *Identity and Individuation* (New York: NYU Press, 1971).
[3] 'Rigid Designation', p. 365.

S2, but we need not identify this ship with b. If we deprive Chandler of the assumption that $b=b^*$, his argument is rendered invalid.

A proponent of Kripke's principle has the best of reasons for denying that $b=b^*$. In S2 b^*'s proper name designates a, since $a=b^*$. Thus, according to Kripke's principle, b^*'s name must designate a in any world (story) in which a exists, and hence in S1. Since b is distinct from a in S1, and b^*'s name designates a, it cannot also designate b. Hence $b \neq b^*$. Since Kripke's principle leads thus directly to the denial of Chandler's tacit assumption, the principle's proponents can scarcely be expected to accept the assumption without argument. And, indeed, the only compelling argument for the assumption would be a demonstration that S1 and S2 (with 'b^*' substituted for 'b' in S2) cannot be given a coherent and intuitive interpretation unless it is assumed that $b=b^*$.

Such an interpretation, however, can readily be given. There are two ships, a (alias c, alias b^*) and b. In S1, a is renovated and its *former* parts, which are no longer parts of a, are used to construct a second ship, b. In S2, a is dismantled and its parts, which are still parts of a (the dismantled ship), are transported and reassembled; that is, a is reassembled. Due to the unavailability of the necessary parts, i.e., certain parts which are not parts of another ship, b cannot be constructed in S2. The crucial point here is that a's original parts cease to be parts of a, and thus become available for the construction of another ship, if they are replaced so that a continues to exist without them; but when a is merely dismantled for transporting, its original parts remain parts of it and cannot be used to construct another ship according to the same plan. Since the parts from which b^* is constructed are parts of a certain ship (a), while the parts from which b is constructed are not parts of that same ship, it should not be surprising that b and b^* are distinct ships.

Chandler's stories, interpreted in this way, suggest that further clarification of the relation '. . . is a part of ———' is needed if we are to resolve certain problems about the identity of composite entities. They do not undermine the principles against which they were directed.

University of North Carolina at Greensboro   © JOHN L. KING 1978

ACTIONS AND BODILY MOVEMENTS

By JAMES MONTMARQUET

IT was not long ago that Prichard confidently espoused the view that all we ever do is will, i.e. that what we actually do in performing any act, including such overt bodily actions as the raising of an arm, is merely to perform an act of volition.[1] More recently, however, there has been an opposite tendency to identify what we do, say, in *raising* an arm, not with any internal mental act, but with the external bodily movement, the *rising* of one's arm.[2] Or to state this view in full generality: every act of bringing about a bodily movement is identical with the bodily movement in whose bringing about that act consists. I shall argue here that this latter view (which I shall henceforth refer to as 'ID') is false.

Although I have noted two places where Davidson appears to uphold ID, my argument against this view will turn on a consideration which Davidson himself has put forward. He writes:

> A man who raises his arm both intends to do with his body whatever is needed for his arm to go up and knows that he is doing so. And of course the cerebral events and movements of his muscles are just what is needed. So, though the agent may not know the names or location of the relevant muscles, nor even that he has a brain, what he makes happen in his brain and muscles when he moves his arm is, under one natural description, something he intends and knows about ('Agency', p. 12).

Now, Davidson's claim, that when I raise my arm I perform an intentional act of bringing about the events which cause my arm to rise, admits of at least three interpretations: (a) that raising an arm is *identical* with bringing about the events which cause my arm to rise; (b) that the latter is a *proper temporal part* of raising an arm; (c) that raising an arm and bringing about the events which cause my arm to rise are *wholly distinct* acts, both of which occur whenever I raise my arm. Davidson's own view seems to be (a), for he writes of an analogous case:

> Doing something that causes my finger to move does not cause me to move my finger; it *is* moving my finger ('Agency', p. 11).

But if (a) is correct, ID must be false, for (a) implies that arm-raisings

[1] See Prichard's 'Acting, Willing, Desiring' in J. O. Urmson, ed., *Moral Obligation* (Oxford, 1949), esp. pp. 188–190.
[2] Advocates of this view include Donald Davidson in 'The Logical Form of Action Sentences' in N. Rescher, ed., *The Logic of Decision and Action* (Pittsburgh, 1967), p. 116; Irving Thalberg in 'Do We Cause Our Own Actions?', *Analysis*, 1967, p. 199; and Raziel Abelson in 'Doing, Causing and Causing To Do', *Journal of Philosophy*, 1969, p. 188. Davidson, it appears, also advocates this view in 'Agency' in Binkley, Bronaugh and Marras, eds., *Agent, Action and Reason* (Toronto, 1971), p. 23.

are identical with acts which *cause* arm-risings—hence, that they cannot be identical with arm-risings. Likewise, if (b) is right, ID must be false, for (b) implies that arm-raisings include (as proper temporal parts) such events as neuron-firings, events which are not included in the rising of an arm. Finally, (c) *is* consistent with ID, but if (c) is right, it is implausible to think that when I raise my arm, I *do* bring about the events which cause my arm to rise. The reason is this: when a person raises his arm, it is not required (under ordinary circumstances) that he first perform some *other* act (that of making a neuron fire, making a muscle move, or whatever), and then go on to raise his arm. Ordinarily, raising an arm does not involve any such preliminaries. When I raise my arm, I simply proceed to raise it; there is no other act which I must do first, only proceeding afterwards to the actual raising of it.[3] To summarize, then: Davidson's claim is true only if it is interpreted according to (a) or (b); on these interpretations, it implies that ID is false. Thus, to show that ID is false, we must only show that Davidson's argument is a sound one.

That argument can, I think, be outlined as follows:

(1) When I raise my arm, I intend to bring about those events which must occur for my arm to rise.

(2) The occurrence of certain neural events and muscle movements is what must happen in order for my arm to rise; therefore,

(C) When I raise my arm, I perform an act of bringing about those neural events and muscle movements, an act which is intentional under the description, 'bringing about those events which must happen for my arm to rise'.

Clearly, the premises of this argument are true; however, to get from them to the conclusion several suppressed premises must be added. One which seems needed is:

(3) If an agent intends to bring about X and succeeds in bringing about X, he will have performed an act which is intentional under the description, 'bringing about X'.

This, taken together with (1) and (2), will imply that if we do bring about the events which cause our arms to rise, we perform intentional acts of bringing about the neural events and muscle movements which cause our arms to rise, acts which are intentional under the description given in (C). So we seem to need just one more premise:

---

[3] Notice this consideration does not militate against (b), for when one act is *included* in another, it would be wrong to say that the agent first did the one and then proceeded to do the other. If (e.g.)I raise my arm twice, it would be wrong to say that first I raised it and then I raised it twice.

(4) When an agent raises his arm, he brings about those events which cause his arm to rise.

Are these added premises correct? It seems plain that (4) is right, for the neural events and muscle movements in question are events which, holding other factors constant, would not have occurred had the agent not raised his arm. In this quite natural sense they are events brought about by him, events which have occurred through his agency. The first of these added premises, however, is not correct. As Chisholm has pointed out (cf. 'The Descriptive Element in the Concept of Action', *Journal of Philosophy*, 1964, p. 616) I may both intend to, and actually, bring about X, without my act's being intentional under the description 'bringing about X'. This, he points out, will happen if X is not brought about in the *way* I intended. (Suppose, e.g., that I intend to kill John, but kill him unintentionally while testing my gun.)

Here it is important to notice that in such cases, among the various acts by means of which the agent has brought X about, there will be a pair, A and B, such that the agent intended that B should be done in some *other way*, while in fact it was done by doing A.[4] Thus, in the previous example, the agent must have intended that his shooting the victim would be done in some other way than by testing the gun. Let us say, then, that an agent brings about X in *exactly the way* he intends if these conditions are met:

(a) he brings about X,

(b) he intends to bring about X,

(c) there are no acts A and B such that: (i) A and B are among the acts by means of which he brings about X, (ii) B is done by

doing A, and (iii) the agent intended that B should be done by doing some other act.

I think that now premise (3) may be correctly formulated as:

(3') If an agent brings about X in exactly the way he intends, then he performs an act of bringing about X, an act which is intentional under the description, 'bringing about X'.[5]

---

[4] Two minor points. There is a degenerate case in which B would not be an act *by which* X is brought about, but the bringing about of X *itself*. This is the case in which it is the intended result itself which happens in an unexpected way—as when I intend to kill S by shooting him, but kill him by running him down in a car. Also, this talk of the acts by means of which X is brought about needn't be taken at face value. We could, as many might wish, reformulate this as talk of the various *descriptions* of the *one* act, the bringing about of X.

[5] A somewhat similar way of dealing with the Chisholm problem is suggested by Alvin Goldman, *A Theory of Human Action* (Englewood Cliffs, 1970), pp. 55–60.

This states a sufficient but not a necessary condition for an act's being intentional. (Suppose that I intend to kill *S* by shooting him twice, but that in testing the gun I accidentally shoot him once, and then proceed to shoot him intentionally. Here I will have killed him intentionally, though not in the exact way I intended.) Acts of bringing about the events which cause my arm to rise do, however, meet this condition. Premises (4) and (1) guarantee that clauses (a) and (b) will be satisfied. Moreover, when an agent brings about the events which cause his arm to rise, he intends that these should be brought about in the ordinary way, and this is the way that they will be brought about; whatever this "ordinary way" is, it is not something which runs contrary to any intention the agent has—thus, clause (c) is also met. By (3'), then, we perform acts of bringing about the events which cause our arms to rise, acts which are intentional as so described. And this implies, by (2), that in raising an arm, an agent performs an act of bringing about the muscle movements and neural events which cause his arm to rise, an act which is intentional (again) described as 'bringing about the events which cause my arm to rise'. Thus, when properly expanded, Davidson's argument is a sound one, and ID must therefore be false. Acts of bringing about bodily movements are not identical with the bodily movements in whose bringing about they consist. The raising of an arm is not the same event as the rising of an arm.

Roosevelt University, Chicago   © JAMES MONTMARQUET 1978

IDENTITY THEORIES AND THE ARGUMENT FROM EPISTEMIC COUNTERPARTS

By ANDREW WOODFIELD

COLIN McGINN, in 'Anomalous Monism and Kripke's Cartesian Intuitions' (ANALYSIS 37. 2, January 1977), argues that the token-token identity-theory can get round an argument of Kripke's against the type-type theory. If this were so, it would be a finding which piquantly undermined Kripke's suspicion that 'the theorist who wishes to identify various particular mental and physical events will have to face problems fairly similar to those of the type-type theorist' (Saul Kripke, 'Naming and Necessity' in *Semantics of Natural Language* eds D. Davidson and G. Harman. Reidel 1972. p. 341).

However, McGinn's argument does not establish the desired result. One may readily construct a parallel Kripkean objection to the token theory. Identity theorists need not despair, though, for there is an alternative countermove against Kripke's argument on both the token level and the type level.

Kripke claims (i) that if pain is identical with C-fibre stimulation, it is necessarily so; (ii) that this goes against our intuition that such an identity would be contingent; (iii) that in the case of other theoretical identities such as 'heat = molecular motion', which are also necessary if they are true, our feeling that the identity is contingent is explicable, because it could have turned out that something which to us was qualitatively indistinguishable from heat (i.e. was an 'epistemic counterpart' of heat) was not molecular motion; (iv) that no such explanation is possible in this case, since any sensation qualitatively indistinguishable from pain *is* pain.

A similar impression of contingency attaches to token-physicalist claims, such as the claim that McGinn's feeling of pain at noon 17.7.76 is identical to McGinn's C-fibres firing at noon 17.7.76. According to Kripke's reasoning, statements of token-identity are necessary too, if they are true. McGinn contends, however, that with token-identities the intuition of contingency *can* be explained in a Kripkean way in terms of 'epistemic counterparts'. He claims that there *are* epistemic counterparts to that pain-token, namely other possible and indeed actual pain-tokens of the same type. The possibility of pains numerically distinct from, but qualitatively identical with, the pain in question accounts for the feeling that this pain could have turned out to be non-identical with that firing of C-fibres. He holds that 'token mental states are like particular tables: they can be (and be essentially) of a type such that other tokens of that type fail to have properties which they, *qua* tokens, necessarily have' (*op. cit.* p. 80). By '*qua* tokens' McGinn means, I take it, '*qua* particulars', or (perhaps better) 'not *qua* anything at all, but through being the particular things that they are'. To avoid any misunderstanding over terminology I shall drop 'token' and use 'particular', as Kripke does.

The argument hinges ultimately upon what may count as an 'epistemic counterpart' of a particular mental state. The question we are considering is: could the particular pain that McGinn had (call it a) have turned out to be non-identical with that particular firing of C-fibres (call it b)? Identity-theorists may feel inclined to say yes. But here is a Kripke-type argument to show that they should not.

'One can imagine a table which looks and feels just like this wooden table yet which is not made of wood. But one cannot imagine a particular pain which is indistinguishable in one's experience from a ($= b$) yet which is not identical with b. For a pain which is experienced as

indistinguishable from pain *a is* pain *a*. "The trouble is", as Kripke says, "that the notion cf an epistemic situation qualitatively identical to one in which the observer had a sensation S simply *is* one in which the observer had that sensation" (Kripke, p. 339). This goes for particular sensations just as much as for sensation-types.

'To hold that other actual pains are counterparts to *a* is like holding that the other tables in my house which are the same size, shape, colour etc. as the one I am looking at are epistemic counterparts to this table even though they are not in this room, are not being looked at by me, and hence are known by me to be non-identical with this one. But the notion of "epistemic counterpart" should not be so construed. It should be understood (following Kripke, pp. 332–3) in terms of an observer's being *qualitatively in the same epistemic situation with respect to all the evidence he had at the time*. A counterpart to a thing is supposed to be something which a person in a certain epistemic position has no reason to distinguish from that thing. But there is no risk of McGinn's confusing other pains with the pain he had at noon 17.7.76 so long as he experiences those others as occurring at different times from that one. In order for a pain to be a genuine epistemic counterpart to *a*, it has to feel the same as *a and* has to occupy the same position as *a* in relation to McGinn's other experiences around noon that day. But any pain that satisfied those two requirements would be *a*.'

This argument assumes that a principle of the identity of epistemic counterparts holds for particular pains but not for particular material objects. That is, it assumes that indiscernible sensation-particulars are identical given that their position in a series of particular experiences provides a basis for phenomenal discernibility. The assumption could be questioned, and perhaps *must* be questioned by a materialist, but it does seem initially quite plausible.

The application of such a principle to *brain-states* would be very implausible, however. Can we not then counter the argument in another way? Let us focus on the *b* side of the materialist identification. Could that firing of C-fibres have turned out to be non-identical with pain *a*? Perhaps not. But at least our temptation to regard the identity as contingent can be explained now that we are looking at it this way round. Suppose that the identification of what was going on in McGinn's nervous system on that occasion had been carried out by a certain neuro-physiologist *N*. *N*, we are supposing, did in fact record the presence of *b* at noon that day, and he also type-identified *b* as a firing of C-fibres. But there could have been a neural event which *N* was unable to distinguish from *b*, which *N* took to be *b*, yet which was not *b*. Various stories could be told about how such a mistake might arise. For instance, there might have been a second firing of C-fibres after *b*. Because *N*'s measuring equipment broke down at noon and then straightaway righted itself, it

did not record b. It recorded only the second firing. N did not realize that two firings had occurred. And because his clock was wrong, N thought that the recorded firing had occurred at noon. One can understand how N, an identity theorist, might have come to think that the firing he recorded was identical to pain a. One can see also that he would have been wrong. Stories like this show why we are tempted to suppose that '$a = b$' makes a contingent claim: it is because we, in stating that $a=b$, might be making the same sort of mistake that N might have made.

Furthermore, a similar account could be given of our intuitions of contingency in respect of type-type identities. Let us concede that if stimulation of C-fibres is pain, it is so necessarily. Nevertheless, there might have been a type of stimulation of fibres (call it shtimulation of D-fibres) in regard to which neurophysiologists were in qualitatively the same epistemic situation as the one they are now in with regard to the stimulation of C-fibres, and which they called 'stimulation of C-fibres'. Because of their primitive neurophysiological theories and crude measuring techniques, many of their detailed psychophysical identity hypotheses would be wrong. In the possible world of D-fibre shtimulation, any epistemic counterpart to pain is pain, but there is an epistemic counterpart to C-fibre stimulation, and it is not pain.

Kripke considered the reverse possibility of pain without C-fibre stimulation, and found that it 'also presents problems for the identity theorists which cannot be resolved by appeal to the analogy of heat and molecular motion' (Kripke, p. 341). Nevertheless, as far as the point about explaining contingency intuitions is concerned, I cannot see that the analogy breaks down. There might have been a phenomenon which scientists confused with molecular motion yet which, because it was not molecular motion, was not heat. Of course, both speculations are far-fetched; they are logical possibilities only. But I find the idea of a neurophysiological misidentification *less* far-fetched than the idea that a confusion of such magnitude might run through physics.

University of Bristol.   © ANDREW WOODFIELD 1978

REPLY TO WOODFIELD

By COLIN McGINN

IN respect of any true a posteriori identity statement containing rigid
designators (or other empirical statement that is necessary if true)
we seem to have an intuition that it might have been false. Elementary
modal reasoning requires that, if we are to insist upon the identity, we
must somehow explain the intuition away as illusory. Kripke suggests
that, for a variety of cases, a 'counterpart' explanation is in order.[1] But
he denies the availability of such an explanation for contingency intui-
tions in respect of psychophysical identity claims, and so concludes that,
in the absence of any other adequate explanation, identity theses are
shown to be false.[2] Now a counterpart explanation requires that it be
possible for numerically distinct items to stand in the counterpart
relation. That relation holds only between 'qualitatively indistinguish-
able' items; it is thus best viewed as an indiscernibility relation with
respect to a certain class of predicates or properties, sc. 'qualitative'
ones.[3] Generally, items may stand in the counterpart relation and be
numerically distinct; but Kripke claims that, in the case of mental events
and states, qualitative indiscernibility entails identity—so a counterpart
explanation of the Cartesian intuition fails.[4] I argued that this was true
of mental types but false of mental tokens.[5]

Andrew Woodfield objects, in effect, that I made the indiscernibility
class too narrow, including only mental type predicates.[6] Had I included
certain other predicates true of particular mental events, he contends,
then, as with types, indiscernibility would have entailed identity, thus
ruling out a counterpart explanation for tokens. He seems to have in
mind two additional sorts of predicate: relational predicates among
mental events, and predicates of date. I think he cannot have wished to
place much weight on the former, since it seems obvious that *sequences*
of mental events could have counterparts if *individual* mental events
could: but the question of date deserves a reply.

To this objection from date I make three retorts. (i) It is not clear
that we should require indiscernibility as to date. Certainly we should

[1] 'Naming and Necessity', in *Semantics of Natural Language*, eds. D. Davidson and G.
Harman (Boston: Reidel, 1972), esp. p. 332 f.
[2] *Ibid.* pp. 334 f.
[3] The relevant notion of counterpart does not exactly coincide with that defined by
David Lewis (see his 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy*
65 (1968), 113–126). His relation is not an equivalence relation, as the intended stronger
relation of qualitative indistinguishability plausibly should be; also we wish the counterpart
relation to be intra- as well as inter-world.
[4] Kripke, *op. cit.*, p. 339.
[5] In 'Anomalous Monism and Kripke's Cartesian Intuitions', ANALYSIS, January 1977.
[6] 'Identity Theories and the Argument from Epistemic Counterparts', ANALYSIS, this
issue.

not include *everything* we know or believe of the particular in question—
for that would block a counterpart explanation in all cases. (That I
know this lectern is made of wood does not, for Kripke, prevent it
having a counterpart made of ice.) Kripke is inexplicit about which
properties to include, but his treatment of examples strongly suggests
that he intends the counterpart relation to hold between *phenomenologically*
indiscernible items. Since the date of a mental event is plainly not a
phenomenological property of it, difference of date does not preclude
the required qualitative similarity. And this restriction seems reasonable,
since the possibility of a purely phenomenological counterpart is
plausibly sufficient to explain the Cartesian intuition. (ii) But suppose,
generously, that we did include predicates of date. That would not help
Kripke. For he would wish a counterpart explanation of the contin-
gency intuition for certain *non-mental* token events, e.g. 'that flash of
lightning = a certain electrical perturbation'.[7] But these events too are
dated. So either they have no counterparts or mental events do.
Woodfield's is therefore not a reply to me that Kripke could consistently
make. (iii) Even waiving these first two points, the objection just assumes
that 'x's pain at t' is (relative to assignments to 'x' and 't') a rigid desig-
nator. But, though it is plausible that the pain that actually satisfies
that description essentially does, it is implausible that no *other* pain could
(similarly for 'the flash of lightning at t'). Suppose I have two pins a and
b and that a causes a pain in me at t; b could have caused a phenomeno-
logically indistinguishable pain in me at that time; and surely the pains
would be distinct (particular) events. (Compare the case of a flash of
light emanating from a certain torch a at t; a torch b could have produced
a similar flash at t; and they wouldn't be the same flash.) So I think that
even if we include date, as I doubt we should, mental tokens do have
possible counterparts, and the usual explanation goes through.

   A counterpart explanation can also be applied, I suggest, to a
different sort of physicalist thesis discussed by Kripke: that identifying a
person with his body (or brain). Kripke apparently endorses a Cartesian
modal argument purporting to establish a dualism of person and
body; viz., we seem to conceive it possible for the same person to
inhabit a distinct body (or brain) and for a distinct person to be housed
in the same body (or brain).[8] But here too, for all Kripke says, a counter-
part explanation can be given: what we really imagine, and what is
really possible, is a qualitatively similar person inhabiting a distinct body
and a distinct person housed in a qualitatively similar body. In fact,
given that Kripke himself holds that a person essentially originates in a
certain sperm and egg, and that the corresponding contingency intuition

   [7] See Kripke, *op. cit.*, pp. 325–6.
   [8] Cf. esp. 'Identity and Necessity', in *Identity and Individuation*, ed. M. Munitz (New York
U.P. 1971), footnote 19.

should be explained in terms of counterparts, it is hard to see, upon his own principles, how to motivate an asymmetry of attitude toward the two physicalist claims—at least so far as concerns the modal arguments.[9]

I set out to satisfy Kripke on his own terms. Woodfield's favoured explanation of the Cartesian intuitions, if I understand it, does not. He seems to say that brain events—types or tokens—admit of distinct counterparts of some sort, but that mental events do not. I shall confine myself to the following remarks. Either the physical events that he claims we imagine ourselves confusing with the given event are genuine counterparts of it or they are not.[10] If they are, then Woodfield's account is the same, in essentials, as mine, now directed toward the physical term of the identity statement. If they are not, then he is open to the charge of failing to do justice to the content of the Cartesian intuition, because what he offers as counterpart is not strictly qualitatively indistinguishable from the given event. His treatment of tokens seems to fall under the first alternative, and his treatment of types under the second. But the more decisive objection is that Woodfield must also, if he is to silence the Cartesian, explain away the apparent intuition that (say) a pain—token or type—is only contingently a C-fibre stimulation; and this will require provision of some sort of distinct counterpart for the *mental* event—which Woodfield has insisted is impossible.[11] Without this there will be an asymmetry between what is possible in other cases, i.e. qualitatively indiscernible experiences caused or accompanied by distinct items, and what is possible in the case of psychophysical identities.

It seems to me, then, that my account stands intact and that Woodfield's cannot do the work required of it.

University College, London   © COLIN MCGINN 1978

[9] For signs of uneasiness over this tension in his views see the final footnote of 'Naming and Necessity'.

[10] Incidentally, I fail to see why, as Woodfield asserts, a pair of items can be counterparts if and only if it is possible for us to *mistake* one for the other; and I find no evidence that Kripke subscribes to such a definition.

[11] I should note that if one accepts the necessity of supervenience as between mental and physical properties, then it seems that any counterpart to a physical event instantiating a certain mental property will, of necessity, instantiate that same mental property. This suggests that no counterpart explanation of the apparent intuition that that physical event might have lacked the given mental property can be given. In 'Mental States, Natural Kinds and Psychophysical Laws', forthcoming in the *Proc. of the Arist. Soc.*, Supp. Vol. 1978, I argue that the intuition must be rejected on other grounds, and give some assessment of the significance (or lack of it) for physicalism of conceding the non-identity of mental and physical types.

ANIMAL WRONGS

By Stephen R. L. Clark

R . G. FREY'S argument on 'Animal Rights' (ANALYSIS 37.4, 1977, pp. 186-9) amounts to this: we cannot argue from the rights of children and imbeciles to the rights of non-human animals because the former have no rights except in virtue of characteristics that are not shared by the non-human. If babies have rights it is because they are *potentially* rational; if imbeciles (lunatics, the senile) are treated as if they have rights it is because they are so similar to 'us' that we feel squeamish at ill-treating them. His other suggestion, that all human beings have immortal souls and no non-human beings do, I will take seriously when he shows one good reason for thinking this either true or morally relevant.

It is not clear whether he thinks there are sound arguments for the existence of animal rights, though as he describes the argument he is out to rebut, from babies' rights, as the strongest argument for animal-rights, I presume he gives little weight to any other. Nor is it clear what the absence of animal-rights would amount to. Historically, such absence has been affirmed chiefly by those who thought that there was no reason why a morally responsible agent should consider the wants and feelings of the non-human just as such (though there *might* be some reason to avoid cultivating a cruel disposition in oneself).[1] Those who have argued for animal-rights, conversely, have usually intended no more than to say that we ought to consider the wants and feelings of non-human beings, that it was non-human suffering rather than human corruption which was the evil to be avoided. More subtle discriminations than Frey deploys are necessary before it is clear where he or anyone else stands on these issues.

But his argument can be considered on its own. Babies have rights, or could be said to have rights, solely because they are potentially reasoning. What is it to be reasoning or potentially so? Chimpanzees and gorillas, once equipped with American Sign Language, can achieve at least a childish capacity for communication, commentary on their own past actions, verbal expression of decision. On those terms, they are actually, and their relatives are potentially, reasoning: accordingly they have rights.[2] But the stakes may be higher than that. The trouble notoriously is that the higher the standard which must be achieved before a creature counts as 'rational' the fewer human beings can meet the standard even

---

[1] See further Stephen R. L. Clark *The Moral Status of Animals* (Oxford, 1977).
[2] B. T. Gardner & R. A. Gardner 'Two-way communication with an infant chimpanzee', A. Schrier and F. Stollnitz (eds) *Behaviour of non-human primates* vol. VI (London, 1971), pp. 117 ff; M. K. Temerlin *Lucy: growing up human* (London, 1976); *New Scientist* 30 June 1977, p. 759 (Koko, a gorilla); see also L. Williams *Man and Monkey* (London, 1967).

potentially. Stoic principles would have us believe that the only truly rational beings are the wise, and it is not clear how many of us even could be wise.

So we must fall back on the second limb of Frey's argument. Imbeciles and the non-rational are treated as if they have rights not because they really and truly do have rights in virtue of the same features that non-human animals possess but because we are squeamish. If this means that the acknowledgement of them as morally-deserving objects rests with us it is obviously true: they cannot compel us to recognise them, they do not even know (let us suppose) what it is to be recognized as morally-deserving. But this is presumably not Frey's meaning. If we recognise them as morally-deserving it is surely in virtue of the features they share with many non-human animals: they do not, cannot, 'reason', but they do feel, desire, communicate, plan (even if only for the short run). The early Cartesians were surely quite right in their intuition that if animals have feelings we ought not to torment and kill them. To be sure, it is difficult for any of us to remember this, that the very same evil is done in oppressing cows, cats and chimpanzees as is done in oppressing, say, microcephalics: possibly, indeed, a greater evil. But that is a perfectly familiar sort of myopia which it is the office of philosophical argument to correct.

I take it then that Frey's suggestion is that we treat or approve of treating imbeciles with consideration simply because they remind us too much of us. He does not explain how it is that non-human animals fail to remind 'us' (whom?) of our own condition—they are certainly very much more like us than Frey's scornful 'Fido' would suggest. But presumably it is a mere matter of fact that the majority of adults in our society, though they may feel squeamish about tormenting animals, do not feel sufficiently squeamish to stop. But suppose we did not feel squeamish about tormenting, say, microcephalics or brain-damaged orphans? A good many people, indeed, probably do not. And many more, if they thought that some advantage could safely be won for the rest of us by torturing such defective humans, would feel no scruples at conditioning themselves out of their own squeamishness. The merely tender-hearted have been little protection for the non-human in the last century: I see no reason to suppose that squeamishness on its own is much of a barrier against the exploitation of the human weak. Parents, of course, may wish to save their children; but then again, they may not. If *parental* rights are all that is really in question then we shall soon be seeing posters proclaiming 'A Parent's Right to Choose'.

And how would Frey feel about such a society? There would then be no injury to adult squeamishness. No-one would then say that torturing sub-normals was wrong. But I presume that Frey does think such treatment is wrong, and does think that such a society would be in

the wrong. But why? Harm is done to the victims, certainly, but that is only thought a wrong (he suggests) because it afflicts our squeamishness: the only wrong done is the injury to *our* sensibilities. If no such injury is done, because we do not care, then no wrong is done.

Of course it is possible that we ought to care, that we would have done gross injury to our souls by educating ourselves out of our squeamishness. But once again: why? If we should care because they deserve our care, then they do so in virtue of the same features possessed by non-human beings, and we should likewise care about the latter. If we should care about them solely as a sort of spiritual exercise, practice for caring about properly rational beings—rather as we may wish children to care for their toys—it would surely be quite irrational to go on *practising* such care in this rather absurd way instead of actually caring for properly rational beings by sacrificing the subnormal.[3]

Secretly, I suggest, we know that we ought to care for the subnormal precisely because they are subnormal: they are weak, defenceless, at our mercy. They can be hurt, injured, frustrated. We *ought* to consider their wishes and feelings, not because we will be hurt if we don't, but because *they* will be hurt. And the same goes for those creatures like them who are of our kind though not of our species. The very same wrong that we can see we are doing when we injure a sub-normal human being we also do in injuring a good many animals. If we ought not to do the one, we ought not to do the other, for the *descent* of our potential victims has nothing directly to do with their susceptibility to injury.

The argument from imbeciles' to animals' rights is not a purely legalistic one to the effect that it is difficult (but not, of source, impossible, given enough *ad hoc* fudging) to devise principles that protect imbeciles and not chimpanzees. It is a *moral* argument that draws our attention to what is *wrong* in mistreating, say, imbeciles: not that it offends us but that it injures sentient, appetitive, partially communicative beings. So does the torture, imprisonment and slaughter of, say, chimpanzees. If the one is wrong (as it surely is), so is the other, for they are, in moral terms, the very same act.

University of Glasgow   © Stephen R. L. Clark 1978

[3] See also A. Broadie & E. Pybus 'Kant's treatment of animals' *Philosophy* 49. 1974. pp. 375 ff.

MORE ON KIRK AND QUINE ON UNDERDETERMINATION AND INDETERMINACY

By M. C. BRADLEY

IN a recent contribution to this Journal ([6]) Mr. Robert Kirk sets out to meet objections which I had raised (in [1]) to an earlier paper of his ([5]) on one of Quine's reasons for the indeterminacy of translation. There are two sets of considerations raised by Kirk's reply, one relating to general constraints and strategy, the other to the specific details of the argument against Quine. Since the publication of Kirk's first paper, and the composition of my own, however, Quine has published a candid and illuminating reappraisal of some of the crucial points involved ([9]; the equally recent [10] contains some of the same material), and his new claims rather change the picture. In the present paper I do three things. In Section I I discuss Kirk's reply to my earlier objection to his argument, without attempting to assess simultaneously the effect of Quine's new paper (also without trying to bring the terminology into line with Quine's). In Section II I then try to assess the effect of Quine's new theses. In Section III I take up Kirk's objections to remarks of mine on constraints and strategy.

I

In his original paper Kirk addressed himself to the argument given by Quine in [8] in behalf of indeterminacy, the argument from the under-determination of physical theory. Quine had maintained that indeterminacy of translation would set in where underdetermination of theory by observation did, while leaving the latter point subject to various possible decisions. Kirk sought to break Quine's argument by describing a case where theory is assumed to be determined up to a certain point, and translation therefore determinate to the same point, but where translation is evidently determinate beyond the point as well. He supposes that we are translating Martian, and that underdetermination sets in only at the level of theoretical physics. Suppose now we apply ourselves to translating the Martians' physics text.

> We could make a partial translation of the Martian text-book, as follows. We put into English everything except Martian theoretical words; and (as our assumption [viz. of translatability of all except theoretical terms] permits) we give the appropriate English syntax to those sentences in which theoretical and topic neutral expressions occur; but we leave the theoretical expressions in Martian. The result is a book differing from an English text-book only in that its theoretical vocabulary is unfamiliar. Now the theory M presented in this book might be isomorphic in the

following sense with the English-speaking physicists' theory A: a theory presented in a foreign text-book, partially translated as described above, is in the relevant sense *isomorphic* with physical theory A if and only if the foreign and English theoretical vocabularies can be so mapped on to each other that, by word-for-word substitution according to this mapping, (i) The hitherto untranslated parts of the book yield sentences which fit into the context provided by those already translated in such a way that the result would be accepted by English-speaking physicists as a text-book of theory A; and (ii) The same goes, *mutatis mutandis*, for whatever supplementary explanations of the theory presented in the book the foreign physicists may be *disposed* to give.

If the Martian theory M was in fact isomorphic with A in the sense explained, then the *only* difference between M and A would be one of vocabulary. But in that case M and A would be one and the same theory (except in a trivial sense). ([5], p. 199)

Certain reservations aside (they are taken up again in III below) I objected to Kirk that

... the theory M in the *Gedankenexperiment* might indeed be isomorphic with A. The trouble is that nothing has been done to rule out the existence of other mappings, compatible with all speech dispositions but not isomorphic in Kirk's sense, and the problem remains of deciding between the various manuals of translation generated by these further mappings. ([1], p. 19)

Kirk's reply to this argument of mine is that I insufficiently appreciated that any two theories (sets of sentences of English) offered in translation of the theoretical sentences endorsed by the Martians 'must be supposed to differ in some non-trivial sense' ([6], p. 139). A trivial difference, according to Kirk, would exist where the theories differed only by a systematic interchange of terms, e.g. of 'electron' for 'neutron'. He holds that *if* an isomorphism in his sense between Martian and English-language physics should exist, then the *only* difference between the Martian theory and the English-language theory would be one of vocabulary. Under such conditions we would be entitled to translate their theory by our theory, and there would be no ground for asserting any indeterminacy beyond the usual inductive variety.

For the obvious reason given in II below, first paragraph, I took it that for Quine *some* theories got by interchanging predicates would differ non-trivially, in a way I will now illustrate. Kirk's picture is as follows. We are confronted with what we can somehow take as Martian physics. Since we suppose that their whole logico-mathematical apparatus is already translated, then the residual problem (supposing we ignore the possibility of singular terms) is merely that of translating some bundle of predicates. Thus we can reasonably picture the task as that of translating some sentences of the form

$$Q_1 Q_2 \ldots Q_n (K x_1 x_2 \ldots x_n)$$

where the Q_i are (already translated) quantifiers, universal or existential, and $K x_1 x_2 \ldots x_n$ is a truth function (truth-functional operators also being already translated) of open sentences whose variable are $x_1, x_2, \ldots,$ x_n and whose predicates are drawn from the so far untranslated set of Martian terms $\{P_1, P_2, \ldots, P_k\}$. Kirks holds that the translation problem is therefore soluble; if we should happen to be able to pair each of $P_1, P_2, \ldots, P_k$ with some predicates $PE_1, PE_2, \ldots PE_k$ of English-language physics in such a way that each of the sentences

$$Q_1 Q_2 \ldots Q_n (K x_1 x_2 \ldots x_n)$$

becomes a truth of *our* physics upon substitution, for each i, of PE_i for P_i throughout, then the only alternative translations would be trivial alternatives or variants in the sense explained.

The difficulty is that even on the assumptions Kirk is making, it cannot be taken to be the case that the only alternatives are trivial ones. To see this, consider the following case (based on an example of Carnap's, [4], p. 61). Suppose we have got a Martian sentence to the point where we have the following translation of it, except for the term whose place is indicated by a dash:

The value of the magnitude M at the space-time point
$< x_1, x_2, x_3, x_4 >$ is —.

(Perhaps the term represented by 'M' is also theoretical; the example is certainly more plausible if it is. I now assume that it is, but ignore the ensuing complications, since they do not affect the point.) Now consider in place of the dash, firstly the term 'rational', secondly the term 'irrational'. Not even the totality of all possible observational evidence could decide either for or against the thesis that the value of M at the place in question is rational; yet, equally (and correlatively) it could not decide either for or against the thesis that the value is irrational. Suppose now that our science says that the value is irrational (because, say, it is a rational multiple of π); then it may happen that the various requirements of Kirkian isomorphism are satisfied so that the problematic Martian term is mapped onto 'irrational'.

My point against Kirk was, broadly, that he fails to recognize the bearing of *other* English-language theories as translations of the Martian, and that the problem remains of deciding between the various manuals of translation generated by these alternative mappings. Thus consider now the different policy of translating the Martian term under discussion as 'rational'. So translating as to put 'rational' in place of the

dash produces a theory not consistent with ours. If therefore there is a mapping of the other P_i into English predicates such that the Martian theory is fully translated, and the resulting theory is observationally indistinguishable from ours, then we have two theories as possible translations of the Martian system, but they are *inconsistent*. Thus one is not a trivial variant on the other.

Now there are a lot of 'mays' in this argument, and I have no wish to excuse them. What I want to make clear is that my original point against Kirk stands; for we see that there may (!) be theories not isomorphic with Martian physics, which yet differ non-trivially from an isomorphic theory (though an isomorphic theory does have trivial variants), and my contention was that Kirk has not met Quine's argument until he has found some way of ruling out those alternative translations.

My illustration is rather rough, for if we suppose, as we do, that the logico-mathematical vocabulary is already determinately translated, then the Martian words for 'rational' and 'irrational' are perhaps already determinately translated. But to object to it on those grounds would not do. Perhaps the Martians, like the Greeks, are primarily geometers, and mark the distinction between the rational and the irrational only at the level of physical geometry, and hence of theoretical physics. But, anyway, we can see the point in a more general way. The point we want is that there may be a physical theory T_1, inconsistent with our theory T, indistinguishable from T by observation, yet differing from T only by a systematic replacement of some or all of the predicates of T by other predicates. It is easy enough to see how this is a possible case without so specific an example as the preceding. Suppose there are two predicates of English which we represent by 'F' and 'G', and suppose the case where T (our theory) includes '$(x)(Fx)$' and T_1 includes '$(x)(Gx)$'. Suppose further that T and T_1 are observationally indistinguishable. Now perhaps on analysis it proves that 'Gx' is to be taken as '$(Hx \supset \sim Fx)$ & Hx'. Then T_1 is inconsistent with T. Now there may be a mapping of a Martian predicate P_j onto 'F', and so on for the other Martian predicates, such that there is an isomorphism between Martian physics and our English-language theory T. But there may also be a mapping of the Martian predicate onto 'G', and a further mapping of the other predicates so that their physical theory is translatable as T_1. The difference between T_1 and T is not trivial; they are inconsistent theories.

But, it may be said to this last claim, we could not suppose the Martian predicate to correspond to 'G', for 'Gx' has a truth-functional structure, and that fact would be known to us since we suppose that we have already dealt (determinately) with the Martian truth-functions. But such a reply is not available. The Martian predicate may be to us, at the translational level we have reached, just another new phoneme sequence exhibiting no structure whatever in terms of previously

isolated and translated morphemes. All that we can say—if we can—
is that that Martian predicate can be rendered by 'Gx', that is, by
'$(Hx \supset \sim Fx)$ & Hx'. We would have no title for saying that if the
Martian term is so translatable, then it must after all be, for the Martians,
a truth-function of other sentences.

II

The preceding line of criticism of Kirk, however, has been somewhat
overtaken by Quine's [9]. In this paper, confining himself to English-
language theories, Quine proposes a method of individuating theories
on which, if there is *any* way of reconstruing the predicates of one conjunc-
tion of axioms so as to turn it into another (empirically equivalent)
conjunction of axioms, or into a sentence equivalent to that other con-
junction, then the theory determined by the first conjunctive axiom is
identical with that determined by the second ([9], pp. 320-21). Suppose
we bring translation into the picture. In his original paper ([8]) on the
underdetermination argument, Quine claimed that the predicates of a
single alien theory T could, consistently with the totality of translational
evidence, be mapped into those of each of two English-language theories,
A and B, such that A and B were inconsistent one with the other. (Just
conceivably the claim was for a mapping of whole sentences, not
predicates. This case is taken up below.) But if such mappings from the
other language exist then there is a mapping of the predicates of A into
those of B (and vice-versa) under which A and B determine the *same*
theory, by the new Quinean test for sameness of theory. But then it
follows that Quine's original argument for indeterminacy of translation
from underdetermination of theory lapses, for now the existence of the
intra-English mappings mediated by the translational mappings precludes
us from getting the point required for indeterminacy, namely that A and
B determine two *different*, indeed *incompatible* translations of T. They
cannot, for now A and B *determine the same theory*. It was only by attribut-
ing to Quine the view that sameness of theory within the translator's
language was not guaranteed by some one—any one—reconstrual of
predicates that his underdetermination argument even *appeared* to work,
and it was on the basis of this eminently reasonable attribution that I
raised my objection to Kirk's Martian argument. But Kirk's rather
casual assumption that any mapping of Martian differing from an iso-
morphic one would be a trivial variant on the isomorphic one now
comes to apply as sober truth, at least as an attribution to Quine. And
the existence of other mappings, though it apparently engenders incon-
sistent theories (as proposed in I above), cannot do so at all on Quine's
new view. Kirk writes in [6] as though Quine's new account of theories
merely confirms his (Kirk's) earlier argument. But, as will now be clear,

it does no such thing. What it does is to re-order things in such a way that Kirk's argument, inapplicable before, now *becomes* applicable.

Kirk states that in correspondence Quine has in fact accepted as a consequence of the new view of theories that the old argument from underdetermination lapses ([6], p. 141, fn. 1). Now Quine is the best authority on what Quine wishes to propose; but even Homer nods, and it seems to me that on the new view of sameness of theories it is not just that the underdetermination argument—very much of an afterthought anyway—is being abandoned; it is that the thesis of indeterminacy of translation itself is being put at risk. What the indeterminacy thesis asserts is that

> . . . manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another. In countless places they will diverge in giving, as their respective translations of a sentence of the one language, sentences of the other language which stand to each other in no plausible sort of equivalence however loose. ([7], p. 27)

Now the variety of translation manuals envisaged in Chapter II of *Word and Object* is a far cry from the variety envisaged in I above in connection with the translation of Martian physics. In Kirk's Martian case everything is fixed except a small batch of theoretical predicates, and it is with respect to just such a case as this (confined to English) that Quine has developed his new account of theories. But in the general case described in *Word and Object* the situation is quite otherwise. On the arguments developed there possible alternative manuals will differ not just in the interpretation they lay upon general terms, but right up to the point where they differ in their ways of *parsing* the alien speech periods so as to arrive at terms and other constructions in the first place. In my [2] I have tried to analyse the part that the possibility of radically different reparsings plays in Quine's network of arguments. It is, according to my analysis, a quite fundamental part, sufficient on Quinean principles to give Quine all the indeterminacy he wants and, indeed, far more than he can accommodate. However, the specific point we want here is this. On the doctrine of *Word and Object* there are sundry different mappings from the *sentences* of the alien language onto sentences of the home language, all compatible with the totality of translational evidence. Consider, then, the totality of sentences deemed true by the foreigner. It maps (according to the indeterminacy thesis) into various subsets of the English sentences, each of which therefore maps into the others. But if each such set maps into the others, how can we avoid taking these sets, not as different theories, but merely as reformulations of the same theory? On the new account of theories, Quine has it that if there is a reconstrual of *predicates* which will turn one axiom set into another then they determine the same theory, pleading as sufficient ground the good

sense of such a decision. Yet what possible basis is there for making this decision at the level of predicates, yet not at the level of whole sentences? If there were any fundamental or even natural way of marking off the case of reconstrual of predicates from reconstrual of whole sentences, we could see that there might be some such ground. But according to Quine's views about parsing there could be no such fundamental or natural (or even objective) way. The level of parsing at which we have worked sentences into certain specific structures of quantifiers, connectives and predicates has, for Quine in *Word and Object*, no objective significance whatever. It is just one admissible way of breaking up the stream of noise, and other such constructions of the noise are equally possible. And this is so whether we are thinking of translation from an alien tongue into an English regarded as unproblematic; or thinking of the construal of a fellow-speaker of English; or, finally, thinking of each English speaker's construal of *himself*. I cannot therefore see any basis for claiming that a reconstrual of predicates of English establishes identity of theory, yet denying that a reconstrual of whole sentences does, except possibly some technical consideration such as finite axiomatizability. Quine's new account of theory identity in [9] relates to finite axiomatizations, and it is, I suppose, arguable that there is no prospect, or at least no presumption, of finite axiomatization of the whole corpus of someone's beliefs about numbers, wombats, coins, the gods, Milton and neutrons. But I do not think this is a hopeful ground, for Quine explicitly does not wish to press for finite axiomatizability: '. . . one could reasonably extend the notion of theory formulations to apply not just to an expression [i.e. conjunctive sentence] but to a recursive set of expressions' ([9], p. 326).

Yet, unless it *is* denied that two alternative admissible translations of the foreigner's total theory are the same theory, the indeterminacy thesis appears to be false. For if the mappings between sets of English sentences generated by the admissible translations of the alien language *do* establish identity of theory, then the admissible total translations of the alien's theory all express the *same* total theory. The appearance of incompatibility, or of disagreement between translations, vanishes, just as under the reconstrual of predicates, and there is a unique translation of the alien language. This admits, indeed, of various formulations (under various manuals), but these no more constitute *different* translations than the axioms with 'electron' and 'molecule' interchanged express different theories.

Still, it could be protested, this does not really get the point right. What the indeterminacy thesis says is that pairs of single English sentences may be correctly adduced as translations of an alien sentence' where such English sentences 'stand to each other in no plausible sort of equivalence however loose'. What we need to show in order to

threaten indeterminacy is not merely that the two total translations of the alien's *whole body of belief* should be taken by Quine to constitute the same theory, but that the pairs of corresponding single sentences should be taken by him to be at least analogously related, say as the same sub-theory. However, the protest might continue, while Quine may very well accept the view about identity of whole theories, he would certainly reject the required view about the identity of sub-theories, and even (perhaps) the very propriety of the notion of *sub-theory*, as a striking passage in *Word and Object* clearly shows ([7], pp. 78–9).

Now what the passage just referred to claims, and what Quine must anyway maintain in order to preserve the indeterminacy thesis, is that we have no title for taking arbitrary pairs of sentences, connected by a suitable mapping, as expressing or determining the same bit or part of the total theory. But if we find Quine taking as examples of *theory*, sentences which do *not* express full systems of the world—which fall short of axiomatizing *all* our belief—then we shall certainly be entitled to doubt whether he *is* rejecting the idea that arbitrary sentences may express a bit or a part of total theory, in short, a sub-theory. And it is barely conceivable that two such sentences should not be treated as determining or expressing the *same* sub-theory if they are correlated under mappings such as those under discussion. Now in [9] we do find Quine doing just what he should not be doing; taking something short of a full system of the world as a theory. In the course of his discussion he denies that Poincaré's conventionalist thesis about physical geometry yields an example of empirically equivalent theory formulations which cannot be reconciled by a suitable mapping of predicates ([9], p. 322). Now there are indeed mappings of the predicates of one (interpreted) metrical geometry upon those of another, but—the point we want— those geometries are plainly not *whole systems of the world* (they do not tell us the date of Milton's birth, for example), and that is so whatever programmatic hopes for the unity of science we may hold to.

There are other pointers in the same direction. In [9] Quine repeatedly returns to the statement, or the implication, that it is theories as whole 'systems of the world' that he is talking about, that is theories as axiom-atizing the whole of someone's beliefs. But it is questionable how serious he is about this. His test for identity of theory, for example, could not be applied to *any* of the systems actually presented in the litera-ture as *theories*; not to axiomatizations of logic, set-theory, elementary number theory or mechanics, for example, since none of these is in fact a full system of the world (none tells us the date of Milton's birth). We do not even know whether there are *any* theories in Quine's sense, since we do not know whether it is possible to capture all beliefs in a single (recursive) set of axioms, even an infinite set. Depending on the details of the notion of *belief*, indeed, we may even know that there are *no* theories

in Quine's sense, by Gödel's argument. For any proposed full system of the world will have its Gödel sentence—a truth of elementary arithmetic not deducible from the theory formulation nor, *mutatis mutandis,* from any of the equivalent formulations. Thus if that Gödel sentence is counted as one of our beliefs, and a theory encompasses all our beliefs, then there are no theories. I am ready to believe that this question of axiomatizability may be, or may rapidly become, a red herring. But at what point does it happen? A given sentence is only a fragment of a *theory* formulation—as against fully determining a *theory* by itself—when? When the formulation it constitutes is not a full system of the world? But when is that, if not just when some truths escape its deductive net?

(In the interests of brevity I have refrained from detailed reference to other writings in which Quine uses the notion of *theory.* But if these were adduced, the point I have been drawing attention to would be exemplified over and again. I invite the reader, for example, to consult Quine's other main discussion of theories, namely Part II of 'Onto-logical Relativity' ([11]), for evidence that he has not previously con-templated restricting 'theory' to full systems of the world.)

III

The other point of dispute with Kirk is about constraints and strategy. In my earlier paper I observed that it was unclear to me that Kirk's approach to Quine was anyway admissible, in view of other aspects of Quine's system. Kirk retorts ([6], pp. 136–8) that what I say is strictly irrelevant to the question whether the argument from the underdetermination of theory, taken as standing on its own feet, is sound. I quite agree with him, and did not intend otherwise. The point I did intend was as follows. The reader of *Word and Object*—the primary source for Quine's indeterminacy thesis—could well protest against Krik's method of argument that it involves a bundle of assumptions most of which are quite untenable according to the contentions of that work. The one I mentioned was that on the arguments of *Word and Object* quantifiers would appear to involve indeterminacy *at whatever level of theoreticity they occur;* hence we cannot suppose, with Kirk, that the point at which translation turns from determinate to indeterminate can be varied *ad libitum,* since quantification will occur on both sides of such a point (providing the point is above the most elementary level). Such difficulties could easily be multiplied. Now Kirk's complaint is that all this is beside the point, since the only question up is whether the argument from underdetermination, taken as sufficient by itself to establish indeterminacy, and as requiring no support from any other considerations, is sound. But the question I was raising is whether the argument can be so taken. In [8], in part explicitly, in part by implication, Quine represented the case as follows. There is fundamentally just one

argument being advanced for indeterminacy. This however takes various forms; in [8] it is the argument from underdetermination, in *Word and Object*, however, it finds different expression. If this position is allowed then, I meant to be saying in my earlier paper, one cannot—as Kirk does—simply isolate the argument from underdetermination as set out in [8], ignoring the various relevant theses of *Word and Object*, for on the doctrine of identity of arguments just noted those various relevant theses will equally be implied by the argument from underdetermination. I have since proposed a detailed analysis of Quine's arguments for indeterminacy ([2]; further relevant material is in [3]), and though my conclusions are favourable to the assumption which Kirk is working on, namely that the argument from underdetermination can and should be considered in isolation from Quine's other arguments, I still think it relevant to point out, as I meant to be doing originally, that Kirk's style of argument is unsatisfactory, unless supplemented by just such an analysis of the interconnections between Quine's various lines of thought.

University of Adelaide   © M. C. BRADLEY 1978

[1] Bradley, M.C.: 'Kirk on Indeterminacy of Translation', ANALYSIS 36.1 (1975), pp. 18–22.
[2] Bradley, M.C.: 'Quine's Arguments for the Indeterminacy Thesis', *Australasian Journal of Philosophy* 54 (1976), pp. 24–49.
[3] Bradley, M.C.: 'Mind-Body Problem and Indeterminacy of Translation', *Mind*, July 1977.
[4] Carnap, R.: 'The Methodological Character of Theoretical Concepts', *Minnesota Studies in the Philosophy of Science* 1 (1956), pp. 38–76.
[5] Kirk, Robert: 'Underdetermination of Theory and Indeterminacy of Translation', ANALYSIS 33.6 (1973), pp. 195–201.
[6] Kirk, Robert: 'More on Quine's Reasons for Indeterminacy of Translation', ANALYSIS 37.3 (1977), pp. 136–141.
[7] Quine, W.V.O.: *Word and Object* (Cambridge, Massachusetts: The M.I.T. Press, 1960).
[8] Quine, W.V.O.: 'On the Reasons for Indeterminacy of Translation', *Journal of Philosophy* 67 (1970), pp. 178–183.
[9] Quine, W.V.O.: 'On Empirically Equivalent Systems of the World', *Erkenntnis* 9 (1975), pp. 313–328.
[10] Quine, W.V.O.: 'The Nature of Natural Knowledge' in *Mind and Language*, ed. S. Guttenplan (Oxford: The Clarendon Press, 1975), pp. 67–81.
[11] Quine, W.V.O.: 'Ontological Relativity', Part II, in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), pp. 45–68.

WITTGENSTEIN'S FAIRY TALE

By INGE ACKERMANN, ROBERT ACKERMANN *and* BETTY HENDRICKS

AT *Tractatus* 4.014c Wittgenstein alludes to a fairy tale involving two youths, their horses, and their lilies in order to illustrate how apparently different things can in some sense stand in an internal relationship of identity. Although most of the references and illusions in the *Tractatus* have been tracked down, none of the extant commentaries

seems to have explicitly identified this story. Its mention by Wittgen-
stein is usually not even noted when 4.014 is discussed, possibly out of
embarrassment that a philosophical point should be illustrated in this
manner. Wittgenstein seems to think that the fairy tale would be familiar
to his German readers, but native Germans we have discussed this
reference with have not recognized the story from Wittgenstein's brief
remark.

In connection with a project on fairy tale symbolism, we have
recently encountered a fairy tale which seems to have just the right
properties to make it the outstanding candidate for the source of
Wittgenstein's reference. This story is included in the well-known 1829
collection of German tales by the brothers Grimm under the title 'Die
Goldkinder'. While several other similar European folk tales and fairy
stories exist, this is the only one to contain precisely two boys, their
horses, *and* their lilies. It appears to slip through most of the thematic
indices to German stories that might be consulted by someone tracing
the *Tractatus* reference because Wittgenstein does not mention the fact
that the boys, and horses, and the lilies are all *golden*. This fact, of course,
is in and of itself irrelevant to his purpose. 'Die Goldkinder' is available
in various English translations of Grimm fairy tales under the title
'The Golden Lads'. Andrew Lang's *The Green Fairy Book* (various
editions, but the Dover paperback (1965) has it on p. 311) contains this
story.

The story of the golden lads is too complex and episodic to relate in
detail, but the relevant aspects can be set forth as follows. A poor man
who repeatedly catches a golden fish is finally instructed by the fish to
divide the fish into six pieces, to feed two to his wife, two to his horse,
and to plant two in the ground. After an appropriate interval, the wife
of the poor man gives birth to golden twin boys, his horse gives birth
to two golden foals, and he has two golden lilies growing in his garden.
At least the lilies and the boys are in some sense in a relationship of
identity because when the one lad meets misfortune by being turned into
stone, his lily simply droops in the garden, and this enables the other
golden lad to recognize his brother's danger, and to ride off on his golden
horse and rescue him. When he returns home, it is discovered that his
brother's golden lily reared up and burst into blossom at the moment of
rescue. In a sense, then, a relationship of identity can be seen between
the golden items in this story even though there are no visible physical
connexions between them. Although this story cannot explain internal
relationships, it seems an interesting illustration of Wittgenstein's
point and it is clear why he might have thought of it in this context.

© Inge Ackermann, Robert Ackermann and Betty Hendricks
                                                                    1978
The University of Massachusetts at Amherst

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. Kemp, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £5.00 (inland), £6.00 (overseas), $15.00 (U.S.A. and Canada); for individuals £4.00 (inland), £4.80 (overseas), $12.00 (U.S.A. and Canada). Each volume comprises four numbers, three of 48 pages and one of 64 pages, appearing within the academic year—in October, January, March and June. Orders should be sent to Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, or placed with any bookseller.

CONTRIBUTIONS. **Articles submitted for publication should be addressed to Mr. Christopher Kirwan, Exeter College, Oxford OX1 3DP.** Contributors are asked to note the following requirements.

Articles should **normally** not exceed 3,000 words in length. **Occasionally**, however, longer contributions can be accepted;

They must be type written in double spacing on one side of the paper only;

Footnotes should be kept to a minimum and wherever possible avoided altogether;

Single quotation marks should normally be used, except for purposes of internal quotation and "scare" quotes.

Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

It is regretted that owing to increased postal charges it is no longer possible to return typescripts unless the following instructions are followed:

Contributors in the United Kingdom should enclose a stamped addressed envelope of suitable size; if immediate acknowledgement is required, a stamped postcard should also be enclosed.

Overseas contributors who wish to have their MSS. returned should send an envelope **and international reply coupons** of the requisite value, whether for air or surface mail.

Galley proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts will be retained by the Editor on the assumption that authors have kept their own copies.

The copyright of articles printed in ANALYSIS remains the property of the author, but contributors are strongly advised, in their own interest, to consult the Editor before consenting to the reprinting of their articles.